Deanship of Graduate Studies
Al-Quds University

Next Generation Sequence (NGS) analysis as a new comprehensive approach for diagnosis of plant microbial diseases

Sameh Naje Ahmad Abu AL-Rob

M.Sc. Thesis

Jerusalem-Palestine

1440/2018

Next generation sequence (NGS) analysis as a new comprehensive approach for diagnosis of plant microbial diseases

Prepared By:

Sameh Naje Ahmad Abu AL-Rob

BSc. Al-Quds University-Palestine

Supervisor: Dr. Ibrahim Abbasi

A thesis submitted in partial fulfillment of requirement for the Master degree in Biochemistry and Molecular Biology / Deanship of Graduate Studies / Al-Quds University

1440/2018

Al – Quds University
Deanship of Graduate Studies
Biochemistry and Molecular Biology

Thesis Approval

Next generation sequence (NGS) analysis as a new comprehensive approach for diagnosis of plant microbial diseases

Prepared by: Sameh Naje Ahmad Abu AL-Rob
Registration No: 21120191

Supervisor: Ibrahim Abbasi, PhD

Master thesis submitted and accepted, Date: 22/ 12/ 2018

Names and Signatures of Examining Committee Members are as follow:

1. Head of committee**:** Dr. Ibrahim Abbasi Signature**:** .....
2. Internal Examiner**:** Dr. Khalid Sawalha Signature**:** .....
3. External Examiner**:** Dr.Abdelmajeed Nasereddin Signature: .....

Jerusalem – Palestine

1440/ 2018

Dedication

To Eman because of her continued support throughout this journey and my whole life thank you

To "Mom", my father's Sole, my brothers and sisters for encouragement and continues support

To Bassel al-Araj, Muhannad Halabi, Mohamad Ali and Ahmed Jarrar soles for their sacrifices

I dedicate this work

S.N

Declaration

I certify that this thesis submitted for the degree of master in biochemistry and molecular biology is the result of my own research accept were otherwise cited, and this thesis is not to be submitted for a higher degree to any other University or institution.

Signature: 

Sameh Naje Ahmad Abu Al-Rob

Date: 22-12-2018

Acknowledgment

I would first like to thank my advisor Dr. Ibrahim Abbasi of the Faculty sciences and Technology / life sciences Dep. at Al-Quds University. His office door was always open whenever I ran into a trouble spot or had a question about my research or writing. He consistently allowed this paper to be my own work, but steered me in the right the direction whenever he thought I needed it. With many thanks Life Sciences Dep. And for the department professors they always grand advises and tools.

I would also like to thank the program Biochemistry and molecular biology professors for the knowledge and scientific tools they grant to us specially program coordinator Dr. Suheir Ereqat she was extremely helpful and facilitating every obstacles, Dr. Rula too was helping and advising through the early stages of my journey.

Special thanks to Dr. Abdelmajeed Nasereddin for his valuable advice in adapting this new technology (NGS), and help in bioinformatics analysis.

This research was in part supported by the Food and Agriculture Organization (FAO): project: *Conducting applied agricultural researches within the research priorities of the National Agricultural Research Centre (NARC)/Ministry of Agriculture (MoA).*

Special thanks to all family members their full encouragements, they were always keeps my spirit high.

Abstract:

Tomato (*Lycopersicon esculentum*) considered as one of the most important required vegetables in worldwide, with a global annual yield of over 100 million tons. Tomato production is affected by different bacterial and fungal infection with a leading bacterial infection caused by *Ralstonia solanacearum* causing a disease known as tomato wilt, in which bacteria invade and extensively colonialize in the vascular tissues blocking water conducting xylem. The main aim of the current study is to screen and to monitor the quantitative abundance of microbiome and fungal organisms in soil and plant parts collected from different tomato green houses. The central methodology of microbiome and fungal evaluation that was used in this study was based on adapting next generation sequencing (NGS) or what is called high throughput DNA sequencing method. This method is relatively new technology that allows mass sequencing and enables the production of a vast array of genomic information from many organisms in parallel and it is provides a separate quantitative counting measurement for each sequenced DNA segment type. Universal primers that amplify the 16S rRNA gene for bacterial species and the internal transcribed spacer ITS region of fungal pathogens were used and the product was sequenced by NGS technology.

The study was performed after collection of tomato plants and soil samples from 7 different greenhouses located in Jenin district over a period of four months starting from October 2017 to late December 2018. Over the collection period a total of 6 collection time points were conducted and in each visit 3 plant samples and 3 soil samples were collected from each growing green house. At the end of the samples collection period, a total of 252 of soil and plant samples were collected. For each collected sample DNA extraction was done, followed by microbiome and fungal DNA fragment amplification using specific primers adapted to be used later in Illumina

MiSeq DNA sequence analysis. The total 252 collected samples were pooled according to their samples nature, green house origin and visit time to form the 85 pooled MiSeq DNA library used in NGS analysis. A total of 170 FASTQ files were produced that consists of paired (read 1 and read 2) for each individual sample. All files were uploaded on Galaxy platform program (usegalaxy.org) and quality filtered. A workflow for sequence analysis that was based on sequence length and selection of fungi unique sequences was applied to analyzed samples after joining the relevant read1 and read2 from each specific amplicon.

The specific microbiome species that were identified in this study were considered from plant pathogenic bacteria, most important identified species are *Ralstonia, Erwina, Pseudomonas, Stenotrophomonas,* and *Achromobacter*. These species are of soil origin and causing different diseases in tomato plant, most important is *Ralstonia* bacterial species that cause tomato wilt disease. The used primers for specific fungi identification were less successful and low numbers of reads were obtained. The main plant pathogen fungi that was identified in both soil and plant leaves was *Alternaria tenuissima,* with some other plant pathogenic fungi species such as *Candida sake, Yarrowia lipolytica, Wickerhamiell apararugosa,* and others.

Different evidence were discussed that support the assumption of the soil being a source of infection since many of the identified pathogens are of soil origin and there was a type of association between microbiome finding in plant and soil samples.

تحليل التسلسل الجيني بواسطة تقنية (NGS) كمنهج شامل لتشخيص أمراض النباتات البكتيرية

إعداد: سامح ناجي أحمد أبو الرب

إشراف: د. إبراهيم العباسي

ملخص

تعتبر ثمرة البندورة (*Solanum lycopersicum*) واحدة من أهم الخضروات المطلوبة في العالم، مع ناتج سنوي عالمي يزيد عن 100 مليون طن. حيث يتأثر إنتاج ثمرة البندورة بعدوى ذات أصول بكتيرية وفطرية من أهمها مرض ناتج عن بكتيريا رائدة تسببها (*Ralstonia solanacearum*) يعرف باسم ذبول البندوره. الهدف الرئيسي من الدراسة الحالية هو فحص ومراقبة وفرة الميكروبات والكائنات الدقيقة في التربة والأجزاء النباتية التي تم جمعها من الدفيئات الخاصة بزراعة البندورة. حيث استندت طريقة البحث على تحديد ومتابعة التواجد البكتيري والفطري وذلك بإستخدام طريقة الجيل التالي من تحديد تسلسل الحمض النووي عالي الإنتاجية (NGS). هذه الطريقة هي تقنية جديدة نسبياً تسمح بتحديد شامل لتسلسل للحمض النووي وتمكن من إنتاج مجموعة واسعة من المعلومات الوراثية من العديد من الكائنات الحية بشكل متوازي وتوفر قياسًا كميًا منفصلًا لكل عنصر من أجزاء DNA المتسلسلة. حيث تم استخدام بادئات عامة لزيادة أعداد مقاطع جينات rRNA 16S للأنواع البكتيرية ومنطقة ITS من مسببات الأمراض الفطرية ولاحقاً تم تحديد تسلسل الأحماض النووية لهذه المقاطع بواسطة تكنولوجيا NGS.

تمت هذه الدراسة بعد جمع عينات من نبتة البندوره وعينات تربة من 7 دفيئات مختلفة في مدينة جنين خلال فترة أربعة أشهر (من أكتوبر 2017 إلى أواخر ديسمبر 2017). خلال هذه الفترة تم إجراء ما مجموعه 6 زيارات لجمع العينات حيث تم جمع 3 عينات من النباتات و 3 عينات من التربة في كل زيارة. في نهاية فترة جمع العينات تم جمع 252 عينة من التربة والنباتات. وبعد ذلك تم استخراج الحمض النووي العينات، ومن ثم مضاعفة مقاطع محددة من الجينات خاصة بالبكتيريا وأخرى خاصة بالفطريات. وبعد ذلك تم تحليل تسلسل الحمض النووي بموائمة طرق مستخدمه ضمن إستراتيجيات شركة Illumina MiSeq لتحليل الحمض النووي. لقد تم تجميع 252 عينة وفقا لطبيعة العينات، ورقم الدفيئة ووقت الزيارة لتشكيل 85 عينة شكلت MiSeq library. لاحقاً وبعد التحليل تم إنتاج ما مجموعه 170 ملف FASTQ يتكون من قراءة 1وقراءة 2 لكل عينة فردية. تم تحميل جميع الملفات على برنامج (Galaxy) لتحليل ملفات

FASTQ/NGS. وبعد فحوصات الجودة لهذة القراءات تم تطبيق سير عمل لتحليل التسلسل يستند إلى طول التسلسل واختيار سلاسل فريدة من الفطريات على العينات التي تم تحليلها بعد دمج القراءة 1 مع القراءة 2 ذات الصلة من كل مقطع تم تحليله.

تمكنا في هذه الدراسة من تحديد أنواع الميكروبات المسببة للأمراض النباتية من خلال مقارنة التسلسل الجيني من للعينات الدراسة بالتسلسلات الجينية المرجعية في قاعدة بيانات بنك الجينات، وأهم الأجناس التي تم تحديدها هي (*Ralstonia, Erwina, Pseudomonas,* and *Achromobacter, Stenotrophomonas*). حيث أن هذه الأنواع هي من أصل التربة وتسبب أمراض مختلفة في نبات البندورة، والأكثر أهمية هو الأنواع البكتيرية من جنس *Ralstonia* التي تسبب مرض الذبول. كانت البادئات المستخدمة لتحديد هوية الفطريات أقل نجاحا وتم الحصول على عدد أقل من القراءات. وقد كانت فطريات الممرضه النباتية الرئيسية التي تم تحديدها في كل من التربة والأوراق النباتية هي *Alternaria tenuissima* ، مع بعض الأنواع الأخرى من الفطريات الممرضة للنباتات مثل (*Candida sake, Yarrowia lipolytica, Wickerhamiella pararugosa*)

تمت مناقشة أدلة مختلفة تدعم افتراض أن التربة مصدر للعدوى حيث أن العديد من العوامل الممرضة المحددة هي من أصل التربة وكان هناك نوع من الارتباط بين العثور على المجموع البكتيري في عينات النباتات والتربة.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Full Word
FAO	Food and Agriculture Organization
ITS	Internal Transcribed Spacer
NGS	Next Generation Sequencing
DDW	Double Distilled Water
GH	Green House
PCR	Polymerase Chain Reaction

Chapter one: Introduction and Literature Review

1. Tomato plant pathology:

Tomato: (*Solanum lycopersicum*) is one of the most important economical vegetable crop after potato. An averageannual world production is estimated of about 100 million tons from 3.7 million ha. (FAO. Statistical database, 2014). Tomato is, considered a rapidly growing crop with a growing period of 90 to 150 days, that needs an optimum daily temperature for growth between 18 to 25ºC with night temperatures between 10 and 20ºC, a condition that is widely available in most regions of the world (Cuartero et al., 2006; Ropokis et al., 2018).Temperature variation above 25ºC and below 10ºC, accompanied by high humidity and strong wind, conditions that produces a lower yield. Also, high humidity encourages a higher incidence of pests and diseases resulting in tomato fruit rotting. Dry climates are therefore preferred for tomato production (Cuartero et al., 2006).Tomato can be grown on a wide range of soils but a well-drained, light loam soil with pH of 5 to 7 is preferred. Water logging increases the incidence of diseases bacterial and fungal diseases wilt and effects on fertilizer requirements amount (Elnesr et al., 2015)

1.1. Tomato diseases: Tomato is the world's second most cultivated vegetable that is exposed to many pathogens during its cultivation or post-harvest storage. it is susceptible to more than 200 diseases caused by different species of pathogenic fungi, nematodes, bacteria, and viruses (Bulgarelli D, 2013; Ikeda-Ohtsubo et al., 2018). The following are the main pathogen species that affects tomato crop in the Mediterranean region:

1.1.1. Bacterial pathogens*:* Plant pathogenic bacteria cause many serious diseases of plants throughout the world (Vidhyase karan 2002;), Bacteria as plant pathogens can cause severe economically damaging diseases, ranging from spots, mosaic patterns or pustules on leaves and fruits, or smelly tuber rots to plant death. Some cause hormone-based distortion of leaves and shoots called fasciation, or crown gall, a proliferation of plant cells producing

a swelling at the intersection of stem and soil and on roots. The following is a summary of the most important bacterial plant pathogens: *Clavibacterm ichiganensis, Pseudomonas syringae, Ralstonia solanacearum and Pseudomonas corrugate,* and others.

1.1.2. Fungal pathogens*:* There are thousands of species of plant pathogenic fungi that collectively are responsible for 70% of all known plant diseases. Plant pathogenic fungi are parasites, but not all plant parasitic fungi are pathogens. Some fungi are hidden inside their plant hosts; these are endophytes, defined by their presence inside asymptomatic plants. Another important group of fungi associated with plants is mycorrhizal fungi. Mycorrhiza means 'fungus root', and it refers to a mutually beneficial association (a type of symbiosis) between fungi and plant roots. The following is a summary of the most important fungal plant pathogens *Alternaria alternata, Colletot richumcoccodes, Stemphylium botryosum, Pleosporaherbarum, Phytophtho racapsici, Sclerotinias clerotiorum.*

1.1.3. Nematode pathogens*:* Plant-parasitic nematodes occur in all sizes and shapes. The typical nematode shape is a long and slender worm-like animal, but often the adult animals are swollen and no longer even resemble worms. Plant-parasitic nematodes range from 250 um to 12 mm in length, averaging 1 mm, to about 15-35 um in width. While nematodes may look dramatically different, they all share some common features. The following is a summary of the most important nematode plant pathogens *Noctuapronuba, Helicoverpazea,* Manducaquin, quemaculata*sexta, Pleosporaherbarum, Phytophthoracapsici, Meloidogynespp and Belonolaimus longicaudatu.*

1.1.4. Viral pathogens*:* Plant virus diseases, like diseases caused by other pathogens, appear to be proliferating at ever increasing rates. Scientific and popular media abound with terms such as new, emerging, re-emerging, and threatening human, animal, and plant diseases. Examples Tobacco mosaic virus, Curto virus, Tomato pseudo-curly top virus,

Tomato bushy stunt virus, Tomato mosaic virus, Tomato mottle Gemini virus and Alfalfa mosaic virus.

2. Diagnosis of pathogenic agents in plants:

Pathogen detection is of most important for keeping the plant health and maximizes crop production. Early detection of plant pathogen enables standing on rapid response settings and emergence plan for disease eradication in greenhouses, country borders, natural landscapes, and other mass production facilities (Mendes et al., 2011; Todd et al., 2018). Accurate pathogen detection with high sensitivity test of plant crops is a fundamental measurement in plant disease management and it plays a major role in the economy and food safety. The failure of traditional methods of early detection of plant pathogenic microorganisms at prompt time and adequately on a routine basis has led to the development of culture-independent, highly specific molecular detection and identification techniques (McCartney et al., 2003; Tedersoo et al., 2018)

3. Techniques used for plant pathogen detection:

Several methods and technologies are used in plant pathogen detection, each has different advantages and disadvantages, new methods that are based on adapting new molecular technologies proved to be much effective and give rapid identification and better sensitivity. Future tech should not only be able to rapidly detect known plant pathogens but also to detect new pathogens through library generation (Berg et al., 2016; Tedersoo et al., 2018). The following are the most important used methods:

3.1. Traditional pathogen diagnosis:

3.1.1. Visual pathogen diagnosis: Although it is the oldest but still it can give an alert for emerging diseases, This approach need experience in interpreting visual symptoms of disease, followed by pathogen diagnosis using naked eye or microscopy apparatus for initial pathogen identification (Tsui et al., 2011). Basically this approach may be the

cheapest and simplest in some occasions, it cannot diagnose pathogens before the symptoms are observable, and it can be performed by the daily available farmers. Of course visual examination is never effective in identifying the pathogen causative agent of the symptomatic disease; especially if this is a result of bacterial or viral infections (Ayliffe et al., 2014).

3.1.2. Pathogen cultures: Culturing is another traditional method which usually takes a few days for microbial pathogens and about 1-2 weeks for different fungi. This method is not practical due to the long incubation time, and the need for expertise and sterile culturing equipments and working area. From the other side culturing methods are very important for pathogen isolation and for further research studies.

3.2. Immunological based methods: These methods are mainly depends on the detection of pathogen epitopes using specific known antibodies. In general, immunological techniques are based on specific and high-affinity binding of the target pathogen epitope to a previously commercially produced antibodies (Ray M, 2017). For plant samples, it is crucial to do sample preparation steps that include pathogen separation and concentration to enable effective pathogen detection (Thornton, 2004). Separation methods are involved physical separation techniques using centrifugation, filtration and electrophoresis, these techniques bring to enrichment of the pathogen antigenic (antigen) macromolecules. The main immunological antigen detection that are used in plant pathogen detection are concentrated on viral antigen detection by enzyme linked immunesorbent assay and slide or plate agglutination methods (Nolasco et al., 1993). New commercially available methods that depends on bio-specific antibody-coated paramagnetic particles have been integrated into biosensors to separate or isolate targeted pathogens were also adapted (Rettcher et al., 2015). However, immunological methods and antibody-coated beads has limitations that are related to high cost of beads required for effective detection, and the need to find

antibodies with specific (monoclonal antibodies) and high affinity to ensure sensitive and specific detection (López MM, 2003; Rettcher et al., 2015).

3.3. DNA based detection (*PCR and Sanger DNA sequence analysis*):

DNA based methods enabling detecting pathogen's specific DNA fragments, this can be achieved by DNA probes; which are old techniques that are not practical and requires radioactive DNA labeling. For DNA probe hybridization there is a need for a previous knowledge on the pathogen type and the use of specific probes for each pathogen that makes this methods complicated(Pravi et al., 2015). A recent molecular techniques were established for sensitive and specific detection of many plant pathogens, currently DNA amplification using polymerase chain reaction (PCR) technology followed by either restriction fragment length polymorphism (RFLP) or more specifically DNA sequence analysis of the amplified amplicons are more common (Abd-Elsalam, 2003; Kuzdralinski et al., 2017; Schaad and Schuenzel, 2010).

3.3.1. PCR: is used to detect the presence of a pathogen's (DNA). DNA first have to be extracted from the plant infected parts, which contains both plant and pathogen DNA, this will be followed by PCR amplification using pathogen specific primers to obtain thousands to millions of copies over three temperatures cycling amplification steps that are repeated about 35 times (Bartlett, 2003; O'Sullivan et al., 2003).For this purpose specific thermal resistant enzyme is used named Taq DNA polymerase that can stand the different temperature cycling. Amplified pathogens' DNA fragments have to be visualized on agarose gel electrophoresis (Barnes and White, 2016; Schaad and Schuenzel, 2010). Alternatively, it is possible to perform real time PCR amplification using fluorescence DNA labeling stains (SYBR1 green, or Syto-9) to avoid the agarose gel analysis step and it is possible to use specific incorporated short probes for specific pathogen detection and

amplification (O'Sullivan et al., 2003). PCR method faces several challenges including: sample preparation, DNA extraction, multiplexing to detect several pathogens.

As it was mentioned earlier amplified pathogen specific DNA fragments can be identified using DNA sequencing technology (Nezhad, 2014; Sharma, 2016). For this purpose the traditional Sanger sequencing technique uses di-dNTPs chain termination to determine the sequence of nucleotides in a DNA strand is used (Knief, 2014; Walker and Lorsch, 2013; Xu, 2016). The pathogens common target sequences are include the most repetitive DNA parts such as nuclear intervening transcribed sequences (ITS 1 and ITS2), or the mitochondrial DNA that include cytochrom b (cyto b) or cytochrome oxidaes (COX 1) genes (Begerow et al., 2010; Kemen et al., 2015; Xu, 2016). For bacterial detection the most target genes are the ribosomal 16s rRNA gene DNA sequences (Davidson and Epperson, 2018; Jo et al., 2016). Most of important plant pathogen genes are available in the GenBank (NCBI,GenBank, NIH) and can be compared with the obtained sequences after performing the initial PCR through *Basic Local Alignment Search Tool* (BLAST) analysis to find the most relevant and related pathogen. This approach is very specific and sensitive and is used in many other fields including clinical diagnosis of human and animal infectious diseases (Burks et al., 1991; Burks et al., 1985; Burks and Tomlinson, 1989).

The more advances in DNA sequencing technologies the greater the speed and efficiency of genome sequencing. Analyzing genomes of plant pathogens provides new information about the processes and genes involved in the host colonization and pathogenicity that offers identification of unknown plant pathogens.

3.3.2. DNA based methods: ***Next generation DNA sequencing (NGS):*** Development of next generation sequencing (NGS) technologies have added to our knowledge greater information that exceeded many data obtained by Sanger sequencing, continuous use of this technology is expected to secure the establishment of DNA databases for both human genes

and plant pathogens (Ameur et al., 2018; Maxwell et al., 2018; Roossinck, 2017).Next-generation sequencing is a technology that dramatically facilitated genome sequencing at lower costs for all branches of life sciences (Aly and Sabri, 2015; Liu et al., 2012). NGS emerged in 2005 using commercial Solexa sequencing technology and expanded rapidly to different systems (Liu et al., 2012; Massart et al., 2014). NGS techniques are basically grouped into sequencing by synthesis or sequencing by ligation. In synthesis-based sequencing, after fragmented DNA samples fixation to the flow cell at specific positions, the complementary DNA strand is synthesized by DNA polymerase while a chemical or fluorescent signal resulting from the nucleotide incorporation is detected to identify the sequence (Liu et al., 2012). The main commercial synthesis-based sequencing technologies are Roche 454 pyro-sequencing, Illumina (solid-phase bridge amplification), and Ion Torrent (Abed et al., 2019; Ambardar et al., 2016; Ravi et al., 2018). Roche 454 pyro-sequencing is based on adhesion of single or primed DNA template to a microbead and amplification using emulsion PCR in such a way that each bead is individually placed within a well, subjected to the flow cell and incubated with DNA polymerase, ATP sulfurylase, luciferase, substrates luciferin and adenosine 5′-phosphosulfate (ASP) (Ambardar et al., 2016). Following polymerase, dNTP is incorporated into the DNA strand and produces ATP as the catalyst required for converting luciferin to oxyluciferin in order to emit light. All free ATP and nucleotides are washed and the process is repeated more times until the DNA template is elongated to the desired length (Ambardar et al., 2016). IlluminaNGS sequencing system is considered the most widely used system, solid-phase bridge amplification is used where each end of a DNA template is ligated with adapters.While one end of adapter-conjugated DNA fragment is attached to a substrate, the other end makes a bridge with immobilized primers and generates clusters of identical template in order to enhance the chemiluminescent signal (Ambardar et al., 2016). This

process continues in a cycle in the presence of a mixture of four nucleotides, followed by image capture while each nucleotide is labeled with a different fluoro-phore. This cycle is repeated until the DNA fragment is synthesized to its target length (Maxwell et al., 2018).The basic principle of Ion Torrent system, by sequentially adding nucleotides, the incorporated nucleotide is detected by measuring pH change due to the release of H+ ions (Abed et al., 2019; Chen et al., 2018). Next generation sequencing systems are able to simultaneously read the sequence of millions of short DNA fragments (typically 25-400 bps in length) (Maxwell et al., 2018; Ravi et al., 2018). These platforms enable quick, low cost, and comprehensive sequencing of complex nucleic acid populations with huge impact on medical academic research and in particular crop genomes which consist of several thousand million DNA base pairs (Ambardar et al., 2016; Gabriel et al., 2014).NGS systems vary in terms of the length of each sequence read, the total bases sequenced, and the price of sequence per mega base (Mb). NGS systems have also been used to sequence plant genomes including: genome of *Theobroma cacao,* apple genome, chickpea, and date palm (Ravi et al., 2018). Many other crop genomes and their wild relatives are current being sequenced.

Sequencing the prokaryotic and eukaryotic plant pathogens and detection of microorganisms existing in infected plant tissues is another application of NGS in plant health. Sequences produced by NGS technology from the infected plant include sequences from any pathogens present. However, this sequence also includes the large genome size of the host plant, complicating DNA sequencing of the targeted pathogens. Hence, depending on the pathogen concentration and sample matrix, an extra enrichment step may be necessary to purify the nucleic acids and to reduce the complexity and the cost of sequencing (Merriman et al., 2012). Among various targeted enrichment techniques in plant biology, the standard PCR-based, hybridization-based, and sequence-capture based

techniques have been widely used along with NGS systems. Standard PCR was used as effective enrichment technique with the earliest NGS systems, but due to the high cost and the challenges for integrating with the new high-throughput NGS technologies its application is only limited to the old NGS devices.[146] Along with the standard PCR, the technique of multiple displacement amplification (MDA) was developed using random hexamer primers to amplify a relatively complete genome. This can significantly increase the quantity and chance of pathogen detection (Ledergerber and Dessimoz, 2011; Ravi et al., 2018). In the hybridization-based enrichment technique, when the complex DNA is applied to the array, the target fragments of DNA hybridize the probes while the non-desired regions are washed away. However, the hybridization techniques needs usage of blocking DNA which restrain non-specific DNA binding leading to the capturing of "off-target fragments".[147] Sequence-capture technology is the alternative to the hybridization-based enrichment which re-sequence targeted regions of genome, where the library of repetitive sequences is depleted first, while the target region is secondly enriched.

Objectives:

The main goal of the proposed study is a proof of concept for the usefulness, and cost effective use of next generation sequencing technology in plant pathogen pre-diagnosis and comprehensive identification as a main tool for obtaining optimal plant growth. The specific objectives are:

1. To diagnose and identify different expected plant pathogens found in soil and plant parts over tomato crop production period. This will be done based on different pathogen DNA segments amplification followed by NGS analysis.

2. To identify on real time morphological and DNA NGS analysis the intensity of plant pathogen infections and suggest an optimal type of pesticide treatment.

3. To associate the above finding with crop varieties pest resistant.

Significant of the study:

Next Generation Sequencing (NGS) is relatively new technology that allows mass sequencing of genetic material, and enables the production of a vast array of genomic information from many organisms. Adapting NGS technology reduces the cost of DNA sequencing compared to Sanger traditional method and this by avoiding time-consuming and tedious traditional cloning steps. In NGS sequencing method it is possible to perform millions of sequencing reactions for part of whole genes or reactions that involve different amplified PCR segments. In the current study we are planning to adapt the NGS analysis for acquiring comprehensive information on the abundance of microbiome and fungi pathogenic organisms that affecting tomato crops from different studied green-houses and over its growth period. The relationship of plant leaves microbiome and soil microbiome is discussed.

Chapter Two: Materials and Methods

1. Sample collection:

In this study samples from tomato plants and soil were collected from 7 different green-houses located in Jenin city. Sample collection was started at the early time of beginning oftomato plant culturing; starting from October 2017 to late December 2018. Over the collection period a total of 6 visits were done from the 7 green-houses (Table 1). In each collection visit; three plant samples and three soil samples were collected from each growing green-house. Samples were collected using sterile collection equipments (scissors, forceps, 100 ml plastic collectioncup, gloves). Caution was highly given not to introduce any contamination to the green-house and not to do cross contamination between the collected samples. Regarding soil samples the three samples were collected randomly from different areas of the green-house and each sample was containing about 50 grams of soil. Plant samples as well were collected randomly from different corners of the green-house and each time few leaves were collected from different tomato plantlets. At the end of the samples collection period a total of 252 of soil and plant samples were collected.

Table 1: (a) Sampling location GPS Coordinates, (b) sampling and visiting date

(a)

Location	Vertical	Horizontal
GH-1	35.392111	32.459719,
GH-2	35.391746	32.459719,
GH-3	35.38959	32.462806,
GH-4	35.405814	32.464134,
GH-5	35.407177	32.462967,
GH-6	35.402564	32.455932,
GH-7	35.40413	32.454891,

(b)

Visit #	Date
visit 1	27/10/2017
visit 2	10/11/2017
visit 3	24/11/2017
visit 4	12/12/2017
visit 5	23/12/2017
visit 6	05/01/2018

2. DNA Extraction

DNA was extracted from a total of 168 samples and this after pooling the three soil samples in one larger container. So, the final numbers of samples (168) was including 126 separate triplicate plant samples (7 green-houses, 6 visits, each time 3 samples were collected), and 42 pooled soil samples. DNA extraction for each collected samples was performed using phenol extraction method as indicated below (Bartlett, 2003).

DNA was extracted from leaf samples directly and for soil samples after their short incubation in 50 ml sterile waster followed by suspension centrifugation. Plant leaves and predicated soil samples were incubated in lysis buffer (100 mMTris-HCl, 20 mM EDTA, 1.4M NaCl, 0.2% mercaptoethanol, 2% hexadecytrimethlammonium bromide) containing 200ng Proteinase K (Sigma, St. Louse, USA). The solution was kept at 65^{o}C for 2 hours followed by 1:1 volume phenol extraction (pH 8.0) and then ethanol precipitation using a final concentration of 0.2M NaCl and the addition of 3 volumes 100% cold ethanol. After incubation at -20 $^{\circ}$C for overnight, DNA was recovered by centrifugation at maximum speed at (14,000 rpm) for 10 minutes, and then the precipitated DNA was washed using 70% cold ethanol. The extracted DNA was suspended in 100 µl TE buffer (1mM Tris-HCl, 0.1mM EDTA) and kept at -20^{o}C until further use(Abbasi et al., 2013).

3. Polymerase Chain Reaction (PCR):

All PCR reactions were performed in a total reaction volume of 25µl using ready-mix PCR kit (Syntezza, Jerusalem, Israel). PCR amplification for each indicated system using 20 pmoles of each direct and reverse primers as they are indicated in Table 2. For each PCR reaction 5µl of the extracted DNA from leaves extract or soil extract was added. So, a total of 168 reactions were performed for each PCR system (group 1-3). It is very important to note that, each of the shown direct or reverse primers has an additional oligonucleotide

sequence (known as Illumina adapters) that is needed for next step of index addition and later used as a site for NGS sequence analysis (Table 2). All PCRs that were used in preparing the NGS (Miseq) library from different samples were performed applying the indicated specific primers and their corresponding melting temperature as in (Table 2). The used temperature profile started at 95°C for 15 min, followed by 35 cycles of 95°C for 30s, 30s at the specified annealing temperature, and 72°C for 1 min, and then concluded with an elongation step at 72°C for 10 min.

Tabel2: PCR systems and their indicated primers' sequences that were used in microbiome and fungi amplification from leaves and soil samples.

PCR groups	PCR system	Purpose	Primers*	Tm (°C)	Ref
Group 1	16S rRNA	Microbiomes	**Direct:** CCTACGGGNGGCWGCAG **Reverse:** AGGACTACHVGGGTATCTAATCC	**60**	**(Abbasi et al., 2018)**
Group 2	18S rRNA	Fungi	**SUNS1:** GTAGTCATATGCTTGTCTC **SUNS4:** GTAGTCATATGCTTGTCTC	**57**	**x**
Group 3	18S rRNA	Fungi	**FITS4:** AGATGTGTATAAGAGACAG **FITS5:** GGAAGTAAAAGTCGTAACAAGG	**57**	**x**

*All used primers synthesized using the below forward and reverse overhang adaptor.

- **Forward adaptor:** TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
- **Reverse adaptor:** GTCTCGTGGGCTCGGAGATGTGTATAAGAGAC

4. Agarose Gel Electrophoreses:

PCR products were run on a 1.5% agarose gel (1.5g agarose, 100ml 1X TAE and 10µlEthidium bromide). The 50X TAE electrophoresis running buffer (242g Tris base, 57.1ml glacial acetic acid and 100ml 0.5M EDTA (pH 8.0)). The Gene Ruler 50bp DNA ladder (Thermo Scientific, # SM0371) was used for sizing PCR amplified products.

5. High throughput DNA deep sequencing using Illumina MiSeq platform:

This approach was adapted from the Nextera microbiome MiSeq DNA sequence protocol to be used in detecting microbiome and fungi in plant and soil samples. For this purpose,three different PCR reactions were performed oneach sample as indicated above. Emphasis on using special primers adapted forusein the Illumina MiSeq next generation sequencing (NGS) system. Each primer is composed of two parts: **First part:** Direct and reverse primers that are specifically designed to target the specific DNA to be amplified (1- bacterial 16S rRNA gene, 2- two systems of fungi ITS). **Second part:**Universal 5' - tailed oligonucleotides (complementary region for R1 connected to the direct primer or R2 connected to reverse primer) known as Read 1 and Read 2. These sites will be used later to add sample barcodes (indices) and for DNA sequencing from both sides.

5.1. <u>Preparation of DNA Template:</u> In order to enable identification of individual sequences of different DNAs in pooled samples i.e., soil and plant leaves, dual barcode sequences are added to the ends of all DNA amplicons. These sequences, known as indices, allow for the identification of individual amplicons and their respective sequences for each sample. A unique multiple indices are used for each different soil or leaves samples, all were pooled one sample tube. Figure 1, shows the primers' regions used in this MiSeq analysis. For the addition of these primers and indices two PCRs were carried out: the first one as indicated above to amplify microbiome and fungi species, and the second PCR as it is indicated below was used for the addition of the dual bar code indices.

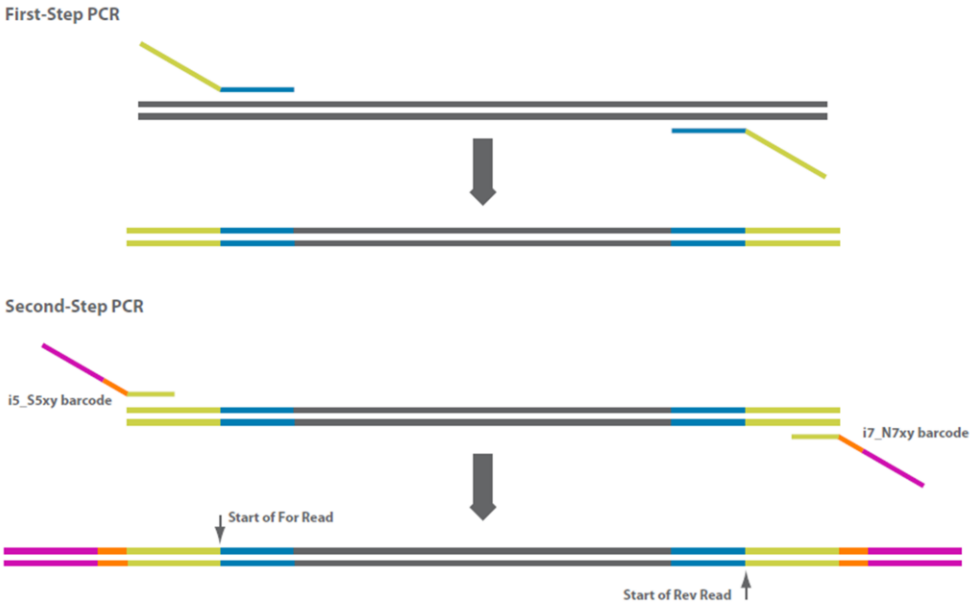


Figure 1. *Overview of the double indexing strategy used in the Illumina two-step protocol. During the Illumina sequencing step the amplified genomic sequence including the specific primers (grey and blue bars) as well as the forward and reverse barcodes (orange bar) are read out.*

5.2. Preparation of MiSeq DNA amplicon library: The library consists of the total amplified 168 samples of both microbiome and fungi samples. The amplified amplicons of the 42 soil samples were treated each as single samples, and the three PCR products were pooled together in order labeled later by their specific indices. The leaves samples (126 samples) were first pooled each three amplified PCR product that are related to the specific green-house in each visit into one sample, so a total of another 42 samples were obtained as a result of pooling each three related samples. At the end a total of 84 samples were obtained; plus extra 3 negative control samples for setting up the background errors. The pooled three amplicons in these 85 samples werw purified using magnetic beads method as indicated below and this was considered as the first step towards the NGS MiSeq library preparation.

5.2.1. PCR products pooling and cleanup using AMPure XP beads (magnetic beads): The importance of this step is to remove un-incorporated nucleotides, primers and salts using highly efficient magnetic beads purification kit (AMPure XP beads kit / Beckman coulter, USA). The following purification protocol was used:

A total of 50 µl of the pooled three amplicons of each soil or leaves samples were transferred into new fresh 200 µl tube (a strip of 8 tubes were used).To each tube containing the pooled PCR products; 30 µl of AMPure XP magnetic beads were added and mixed well.The mixed components in the strips were incubated at room temperature for 5 minutes. This was followed by transfer of the strips containing the PCR products and the beads to 96 well magnetic plate stand, the strips were left for another 5 minutes until the beads attached to tube side. (DNA suppose to bind to the magnetic beads).The solution containing unbound materials (PCR buffers and salts) were removed by gentile pipetting without disturbing the attached beads (for this purpose multichannel pipette was used).The

beads and the bound DNA were washed with 200 µl of freshly prepared 80% ethanol and left for another 1 minute while strips are not in the magnetic plate.Then the strips containing the beads and the washing solution were returned to magnetic plate and left until the beads attached to tube side (about 1 minute), and then the washing ethanol solution was removed by pipetting. The washing step was repeated, and at the end of the second wash strips were left containing the beads without any ethanol on the magnetic plates for about 5 minutes to complete drying. Later the strips were transferred to regular PCR tube strips holder, and then 30 µl of double distilled water were added to elute bound DNA. The beads in DDW were left for about 2-3 minutes. Lastly the strips were transferred to magnetic plate tube and then pipette the eluted DNA (20 µl) into fresh tubes.

Note: the transferred 20 µl suppose to have a representation of the three amplicons (microbiome and two fungi PCRs) for each analysed single sample.

5.2.2. Index addition by PCR: This is the second stage PCR that was used to attach the dual indices (i5 and i7) linked to Illumina sequencing adapters. For this PCR the below indicated indices names written in the sample sheet were used (commonly they are named as: N7XX and S5XX). Index additions were performed in ready-mix PCR kit (Syntezza, Jerusalem, Israel). The following is the composition of each PCR reaction: (reaction total volume= 25 µl).

Quantity	Material
S5 XX index primer 1	5 µl
N7 XX index primer 2	5 µl
DNA (quantity to be transferred directly to this tube at the end of magnetic bead purification).	15 µl
Total	**25 µl**

The PCR was performed on Thermo-cycler using the following program:

- 5 min at 95°C.
- **12 cycles: each composed of**
  - 30 seconds at 95°C.
  - 30 seconds at 55°C.
  - 30 seconds at 72°C.
- A final elongation step at 72°C for 5 min.

Representative samples were analysed by Agarose gel electrophoresis in order to prove the success of dual index addition.

5.2.3. Final preparation of MiSeq pooled and barcoded sequencing library: After addition of the indexes by the second PCR, all the reactions were purified using the AMPure XP magnetic beads protocol, as indicated above and exactly as done after pooling the first PCR. At the end of this purification step all the individual eluted PCRs were pooled into one tube by mixing 10 µl from each eluted PCR product. At this stage the library is ready to be sent for next generation DNA sequencing (NGS) Service Company (sequencing was done on Miseq machine using 500 cycle kit from Illumina Co.).

5.2.4. Library quantification and normalization: It is very essential to quantify all the 86 pooled samples. This quantification is best done by Tapestation electrophoresis system (Agilent Technologies, CA, USA), that analyze DNA fragments (or PCR amplicons) from 35 to 1000 bp in terms of quality, size and concentration.

5.3. Bioinformatics analysis: Raw Illumina sequencing data was generated from all analyzed PCR amplicons as FASTQ files of read1 (forward) and read2 (reverse) for each individual sample. These sequence reads were uploaded to Galaxy platform at (usegalaxy.org) for further sequence processing and analysis (Ref: Afgan et al). Initially raw sequences were filtered for quality control at a phred score of 20 equivalent to 99% confidence of each nucleotide, followed by merging forward and reverse reads, the amplified specific genes were selected according to their specific sequence length and sequence identity. The selected sequence reads from each soil and plant leaves were analyzed for sequence homology above 97% using BLAST analysis tools in order to determine number of reads related to specific microbiome or fungi operational taxonomic unit.

Chapter three: Results

1. PCR amplifications and MiSeq DNA library quality analysis:

1.1. 16s rRNAMicrobiomePCR: Successful amplification was tested each time on 1.5% agarose gel electrophoresis for both microbiome and fungi PCRs. The three used systems are standard systems that are already well established and optimized from many other previous studies. Figure 2 showed a represented agarose gel electrophoresis results of the bacterial 16s rRNA PCR system for both soil and plant extracted DNA. A successful DNA amplification was obtained from all examined 168 soil and plant samples.

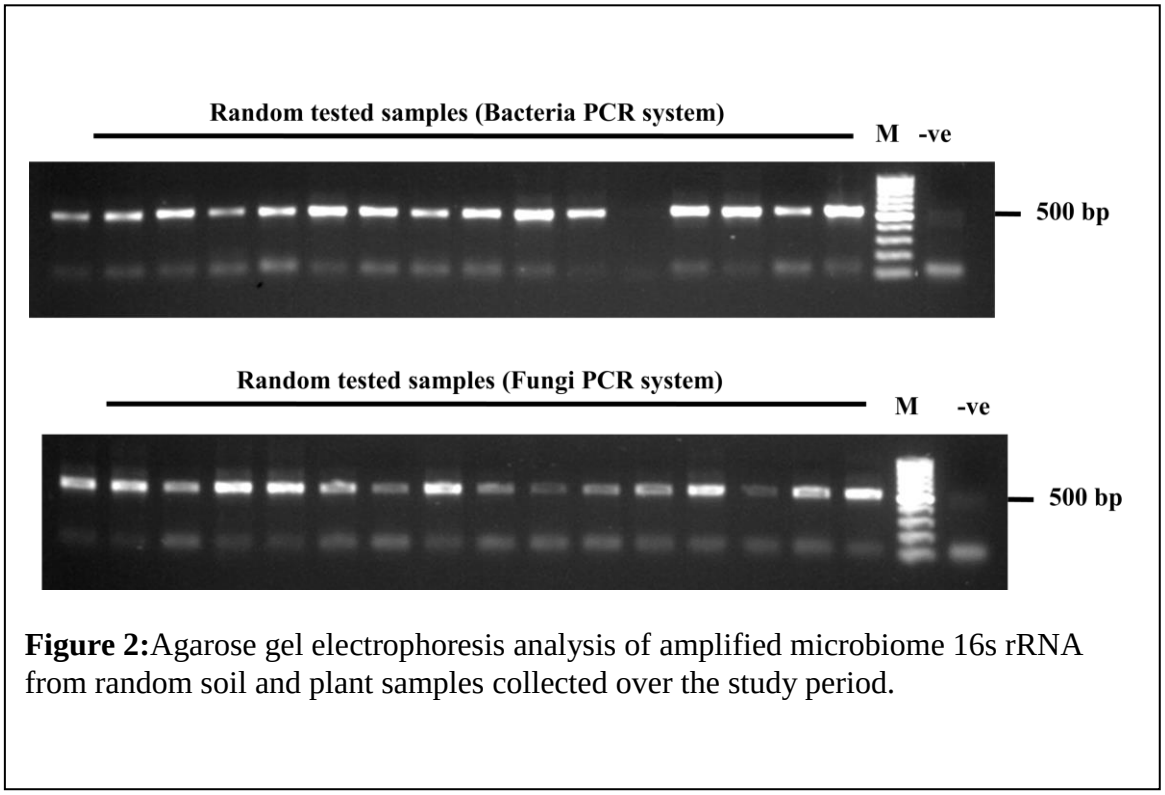


Figure 2: Agarose gel electrophoresis analysis of amplified microbiome 16s rRNA from random soil and plant samples collected over the study period.

1.2. ITS rRNA fungi PCR:

Although the two fungi PCR systems were optimized by other researchers and these primers were used in plant fungi optimization, still not all the tested samples gave a positive PCR amplification as was seen by the microbiome analysis. Only 40 samples from the total 82 samples analyzed by NGS showed a positive fungi PCR amplification. (Figure) 3 shows the results of random samples analyzed on agarose gel electrophoresis applying the

ITS rRNA fungal PCR. The second fungi PCR system was less effective and less number of samples gave a positive PCR amplification. All the obtained amplicons from soil or leaves plant samples whether they showed positive or negative results were included in the MiSeq DNA library.

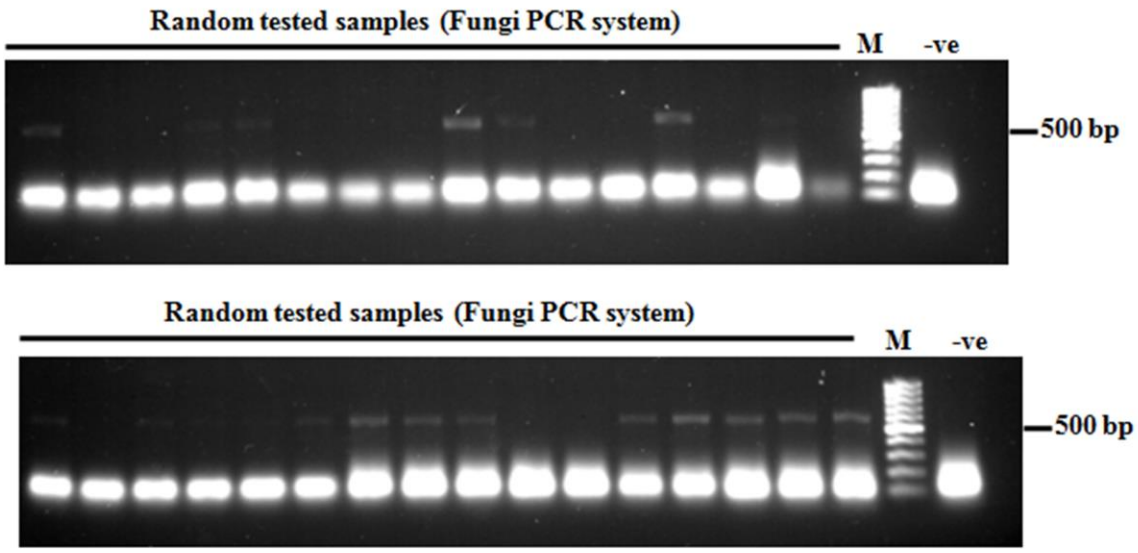


Figure 3: Agarose gel electrophoresis analysis of amplified fungi ITS gene from random soil and plant samples collected over the study period.

1.3. MiSeq DNA library analysis and index addition quality control check: After the second PCR that involves the addition of two indices (S5xx and N5xx), a quality analysis of the obtained pooled and purified DNA amplicons was performed. The indices addition introduces about 130 bp addition on each PCR. So, after the indices addition random samples were chosen to ensure that the index addition PCR was successful. As seen in (Figure 4), for each type of PCR system there was a 130 bp advanced shift in amplicon size as results of indices addition. Later and after the purification of all indexed labeled amplicons and their pooling in one tube to be used in MiSeq Illumina sequences analysis, a Tapestation electrophoresis analysis (Agilent Technologies, CA, USA) was performed using SYPBR green I fluorescence detection system and this in order to analyze the produced library at the quantitative and qualitative level. Figure 5, shows the Tapestation results

2. <u>NGS MiSeq sequence data:</u>

A total of 170 FASTQ files were received from the service company that performed the NGS run. The 170 files are considered the raw Illumina sequecing data and they are consists of two different files: Read 1 and Read 2 for each individual sample (total number of sequenced samples is 85). The total size of the obtained FASTQ files reach up to 778 Mb that contains the sequences for each included amplicons that were produced from each individual analyzed soil or plant leaves samples. All files were uploaded to Galaxy platform program (usegalaxy.org) as indicated in material and methods. A workflow for sequence analysis that was based sequence length selection and selection of fungi unique sequences was applied to all files after joining the relevant read1 and read2 from each specific amplicon

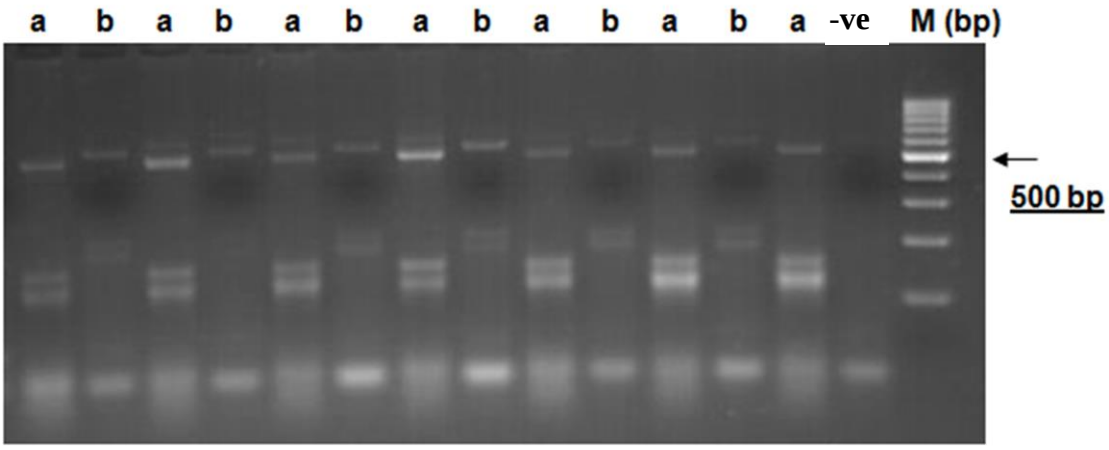
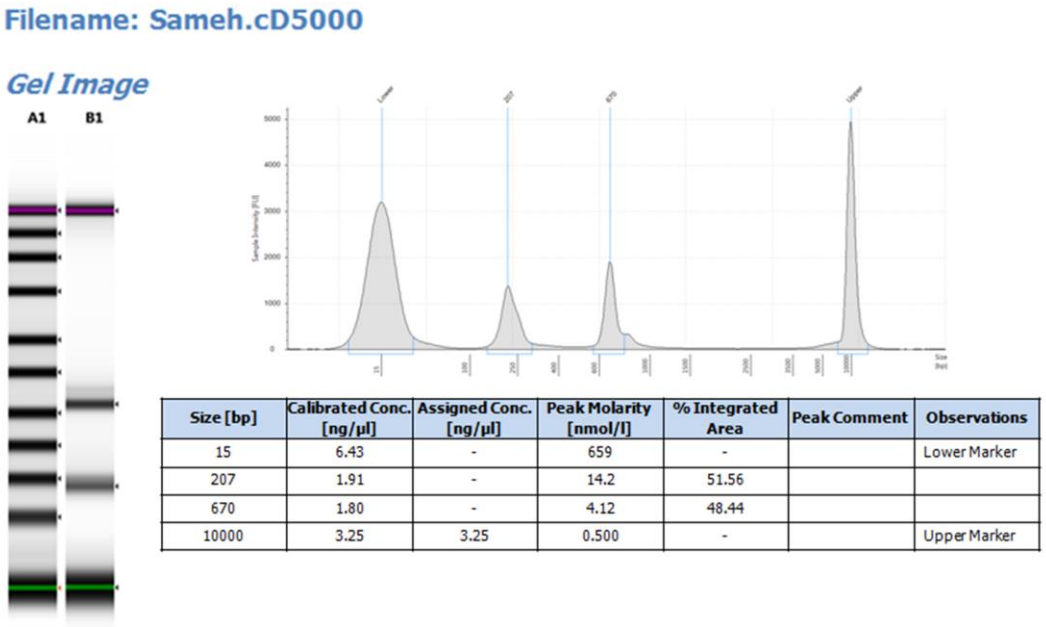


Figure 4: Band size comparative analysis for some selected amplicons used in MiSeq library preparation, before and after indices addition. About 70 bp shift in amplicon size seen in (b: after index addition) and (a: without index).



Size [bp]	Calibrated Conc. [ng/µl]	Assigned Conc. [ng/µl]	Peak Molarity [nmol/l]	% Integrated Area	Peak Comment	Observations
15	6.43	-	659	-		Lower Marker
207	1.91	-	14.2	51.56		
670	1.80	-	4.12	48.44		
10000	3.25	3.25	0.500	-		Upper Marker

Figure 5: MiSeq library quality control analysis using a sensitive fluorescent DNA analysis performed on Tapestation electrophoresis machine. Two bands are seen in size 200 bp (NGS primers) and 600 bp (the amplified bacterial and fungi amplicons before purification).

2.1. NGS Microbiome and fungi data analysis : Using Galaxy workflow it was possible to obtain the exact number of total amplicons for each separate analyzed sample from soil and plant leaves and which belongs to the specific collection date. Figure 6 shows a histogram draw for soil and plant amplicon reads representing bacteria and fungi amplified DNA fragments from all studied green-houses all over the collection period. It is clearly seen that total reads obtained from soil samples were much less than those obtained from plant samples, also total reads in soil started to increase at the end of the study in most of the green-houses. On the other hand, the total reads obtained from plant samples showed great variations among different green-houses and between different visits as well (Figure 6 b). Although in the initial visits green-houses 3 and 4 showed the highest reads all over the study that reached up to 80,000 reads, while green-houses 7 and 6 showed the lowest number of reads all over the study which indicates a cleaner habitat and healthy conditions.

In order to have clear and precious information about the microbiome and fungi abundance in different green-houses all over the visits and to highlight the most important microbiome, a separate analysis of each green-house was performed as indicated below.

2.1.1. NGS microbiome analysis in soil and plant samples (overall observations): For each of the studied green-house the most abundant five microbiome species were determined in each time of sample collection. For this purpose the total reads that belong to the bacterial 16 sRNA were translated to unique operational taxonomic units (OTUs) which are defined as the closely related group, or the similar sequences grouped by DNA sequence similarity of a specific taxonomic marker; which is in this case the 16 sRNAS gene.

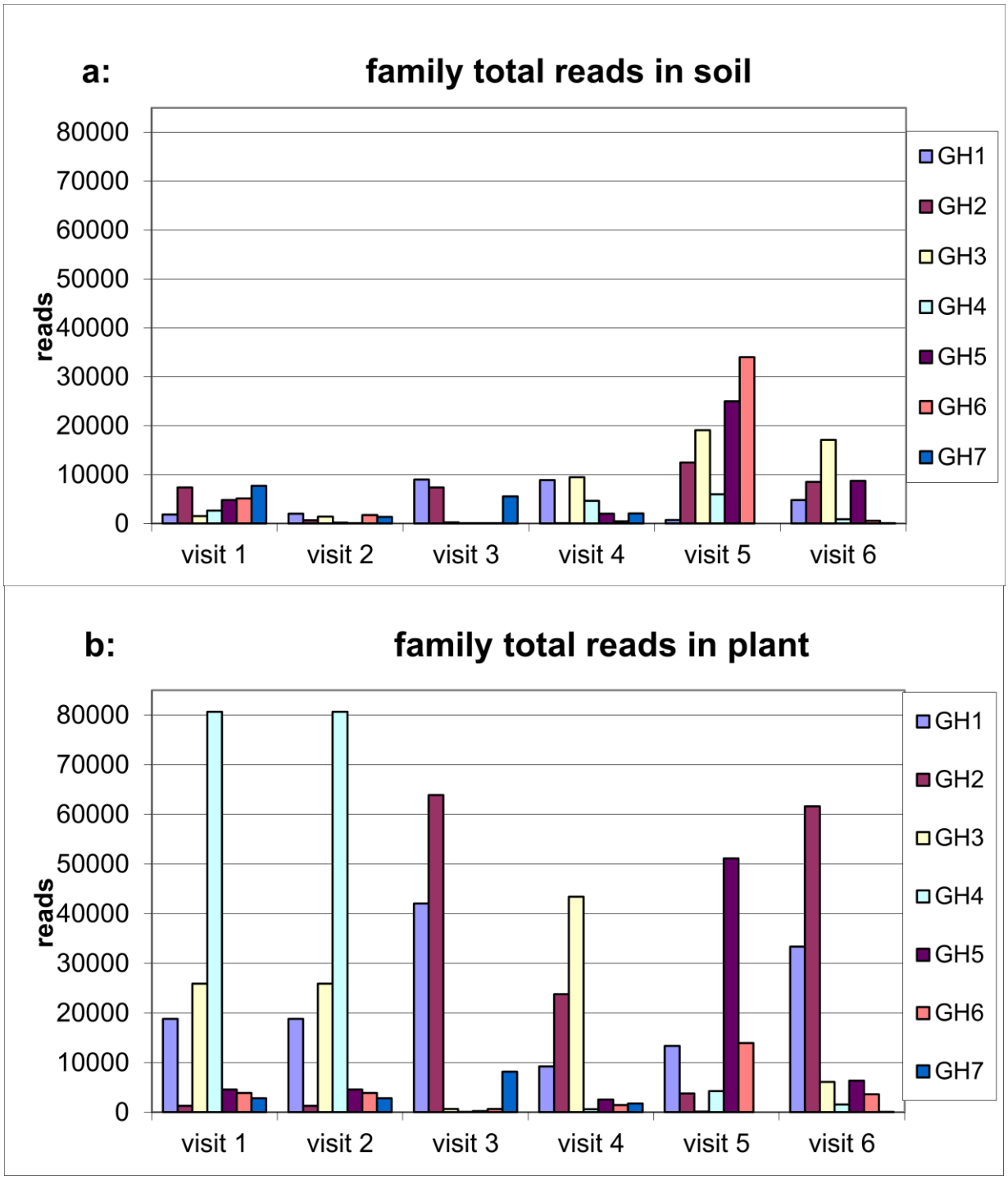


Figure 6: Comparative histogram analysis of total bacterial and fungal reads found in soil or plant leaves obtained per green house at each visit over the study period.

In general there was a slight inverse association between microbiome abundance found in soil compared to that in plant leaves and this was true in GH1, GH2, and GH3 (Figure 7). All over the study period the plant leaves microbiome reads were found to be higher, and during the visits 3 or 4 (middle of the study period) and 6 (at the end of the study period), the microbiome in leaves showed an obvious increase compared to soil in GH1, GH, and GH3. This finding is discussed later with suggestions for such possible associations. Green-houses 4-7 were in total showed less abundance of microbiome with a relative complete consistence in microbiome quantities in soil and plant leaves in GH4-GH7 (Figure 8). Specifically, GH5 showed about level of bacterial contamination in soil and plant leaves all over the collection period (6 visits). Also, it was noted that GH3 and GH4 initially started with higher microbiome content in plant leaves than in soil, that could suggest an external contamination factor originated from the tomato plantlets at the initial time of grow. It is worth to mention at the current time that the water used in irrigation also could be an important factor in origin of the bacterial contamination as discussed later.

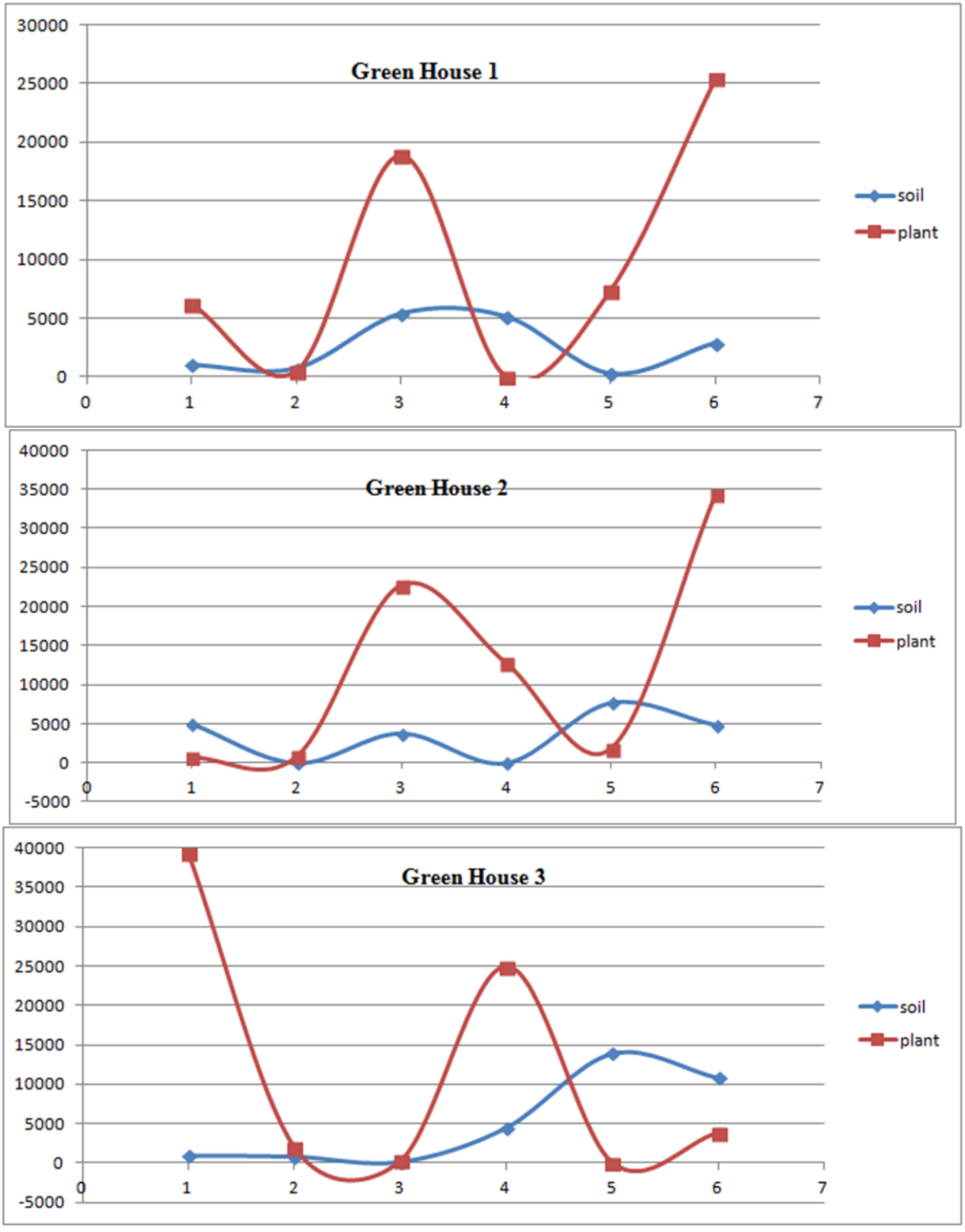


Figure 7: Comparative total microbiome relationship in soil and plant leaves (green houses 1-3).

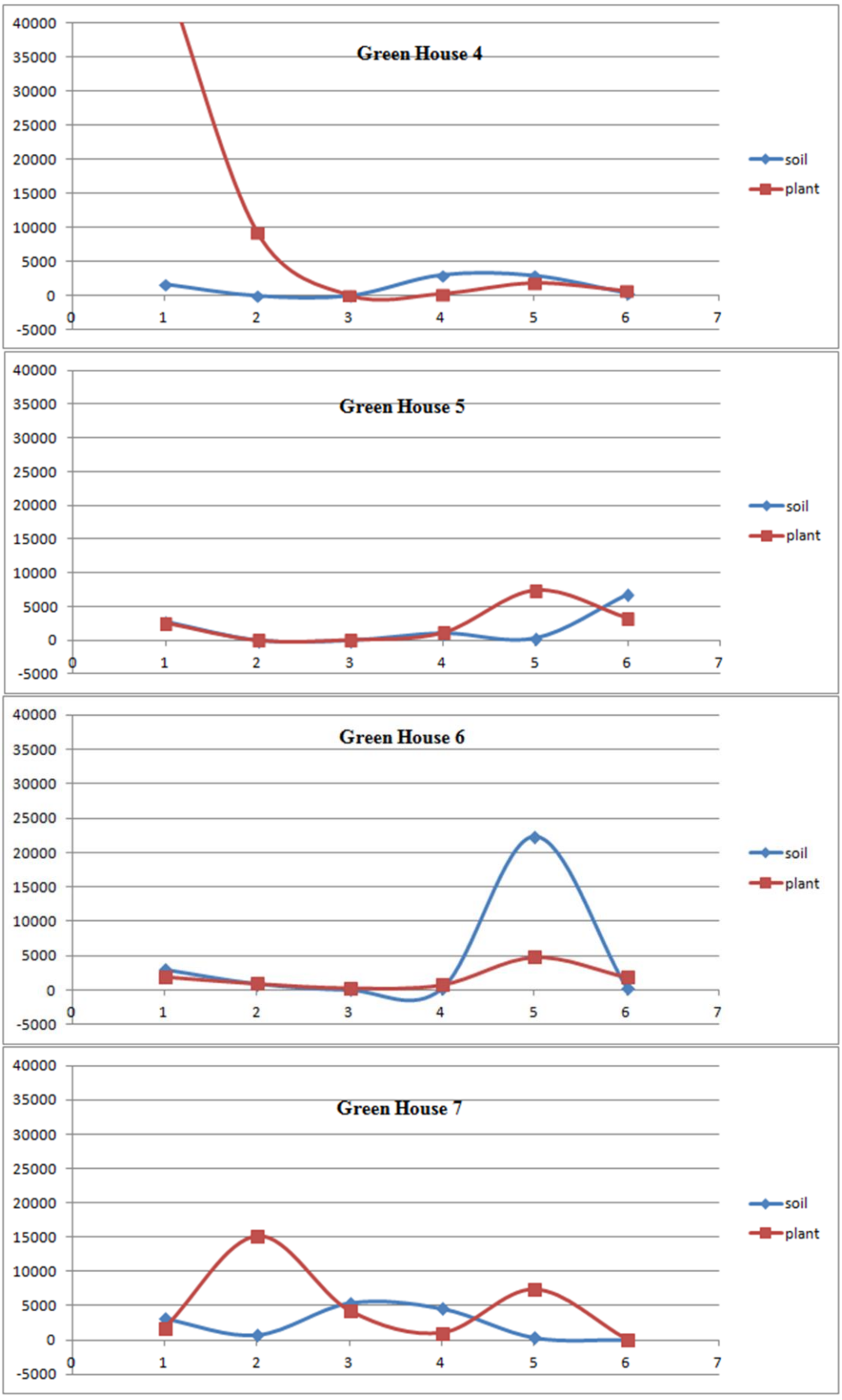


Figure 8: Comparative total microbiome relationship in soil and plant leaves (green houses 4-7).

2.2.2. NGS microbiome analysis in soil and plant samples (species specific observations): Green-house number 1:A highest number of reads were recorded for *Escherichia albertii* and *Ralstonia pickettii* bacterial species during visit 3 in plant leaves (Figure 9). The two indicated bacterial species also were relatively high in soil samples and exactly at the same collection time, this could suggest that the bacterial contamination was originated from the soil. *Ralstonia* bacterial species was clearly noticed to be the most abundant bacterial species at most visits and also in both soil and plant leaves was seen at the same collection time. Another two important bacterial species were detected in a high number of reads are (*Pseudomonas* species and *Erwinia* species). *Erwina mallotivora* in fact recorded the highest number of reads that reached up to 12,149 at the last time of samples collection (visit 6). Analyzed soil samples during visit 4 and plant leaves samples at visit 3 and 4 showed the highest peak of *Pseudomonas* different bacterial species as the predominant species with a number of reads that ranged from 295-2012 in soil samples during visit 4, while number of reads related to *Pseudomonas plecoglossicida* reached up to 4181 in plant leaves at visit 3. This result suggests that *Pseudomonas plecoglossicida* species could be established first on plant leaves and then transferred to soil through leaves falling onto the soil.

An interesting observation was observed related to *Stenotrophomonas geniculata* that was found in 5527 number of reads in plant leaves and in 629 reads in soil samples during visit 6 (Figure 9). The same observation with different number of reads were also observed in GH2, GH3, GH4, and GH7 (Figures 10-13), which suggests the appearance of this type of bacterial species at later stages of tomato growth. These species were not seen in GH5 and GH6 (Figures 14-15) which were much cleaner (at the microbiome level) and were seen to be in healthy environment.

Green-house number 2:The microbiome total reads in GH2 were recorded to be the highest during visits 3-6, and in after visit 6, this green-house was completely eradicated and closed for sterilization (Figure 10). Different bacterial species of *Pseudomonas* started to appear after visit 3 and there was a drastic shift in microbiome number of reads from low to high and this pattern of higher number of reads that reflecting higher number of bacterial species was kept all over the coming visits. The other important bacterial species that were also highly recorded are: *Stenotrophomonas geniculata* and *Pseudomonas plecoglossicida* that were found to have number of reads of 11,891 11,394 respectively, and this was followed by other *Ralstonia* species. This finding could suggests that these three species together with *Erwina* species are the main bacterial species that produce this drastic infection that lead to green-house eradication.

Green-house number 3:The mode of microbiome pattern in GH3 was different from the others in that, originally and from the first visit there was a high number of bacterial reads related to *Denitrobacter permanens* and *Achromobacter arsenitoxydans* found in plant leaves samples but not in soil samples (Figure 11). These bacterial species are supposed to be originated from the tomato plantlets. A similar observation as the case in GH1 and GH2, different *Pseudomonas* species were abundant at visit 4 and in both soil and plant leaves collected samples. The fact that *Pseudomonas* species in plant leaves was less at visit 5 compared to soil samples and after their abundance at visit 4, this supports the previous suggestion of the origin of plant leaves infection by *Pseudomonas* bacterial species of soil origin.

Green-houses 4 and 5:They were the healthiest in their appearance during all the visits of samples collections, and as it was mentioned above they showed the lowest number of microbiome reads (figure 6, figure 12 and Figure 13). The most important observation in

these green-houses are related to the initial microbiome records in GH4 that were found to be extremely high at the first visit (*Achromobacter piechaudii* 22,594 reads and *Denitrobacter permanens* 10,363 reads). This suppose to be at time of tomato plantation, then the microbiome reads dropped to very low levels, that suggest a well treatment and prevention precautions were taken by the owners of this green-house. The second important observation was related to the abundance of different *Pseudomonas* species at visits 3-4 (Figure 13), a consistent result that was also seen in GH1 and GH2 (Figure 9 and Figure 10).

Green-house number 6: also was noticed to be healthy and showed a good plant growth compared to other green-houses, and this fact was easily to be predicted from the low number of microbiome reads all over the collection time (Figure 14).

Green-house number 7: The highest recorded bacterial species was Achromobacter*ar senitoxydans*that reached up to 10,811. In general there were several bacterial species of soil origin recorded in both soil and plant leaves samples, such as *Erwinia soli*, *Rheinheimera soli*, *Escherichia albertii*, and other *Pseudomonas* species.

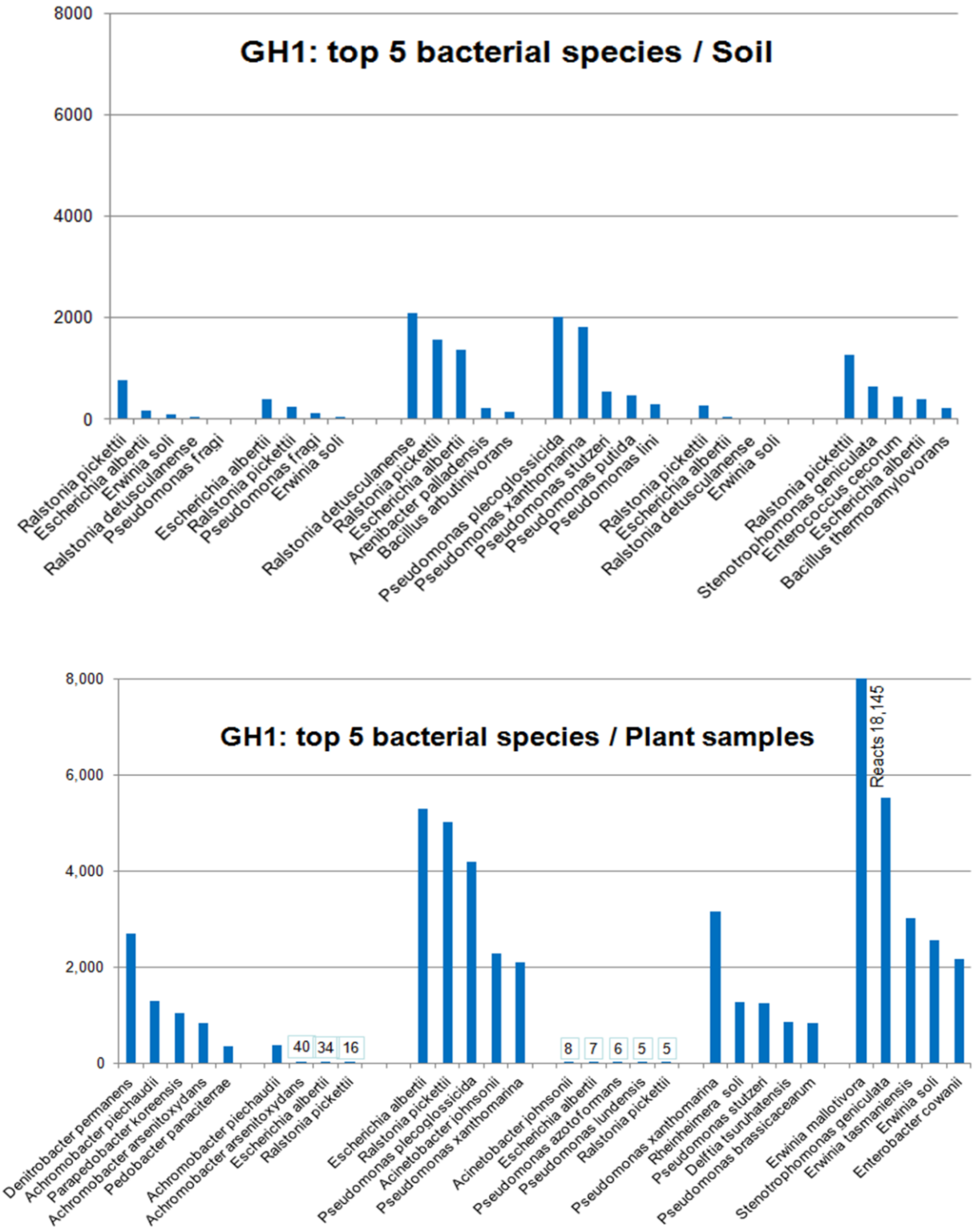


Figure 9: The most abundant five bacterial species obtained in each visit in soil and plant leaves in green house 1 (GH1).

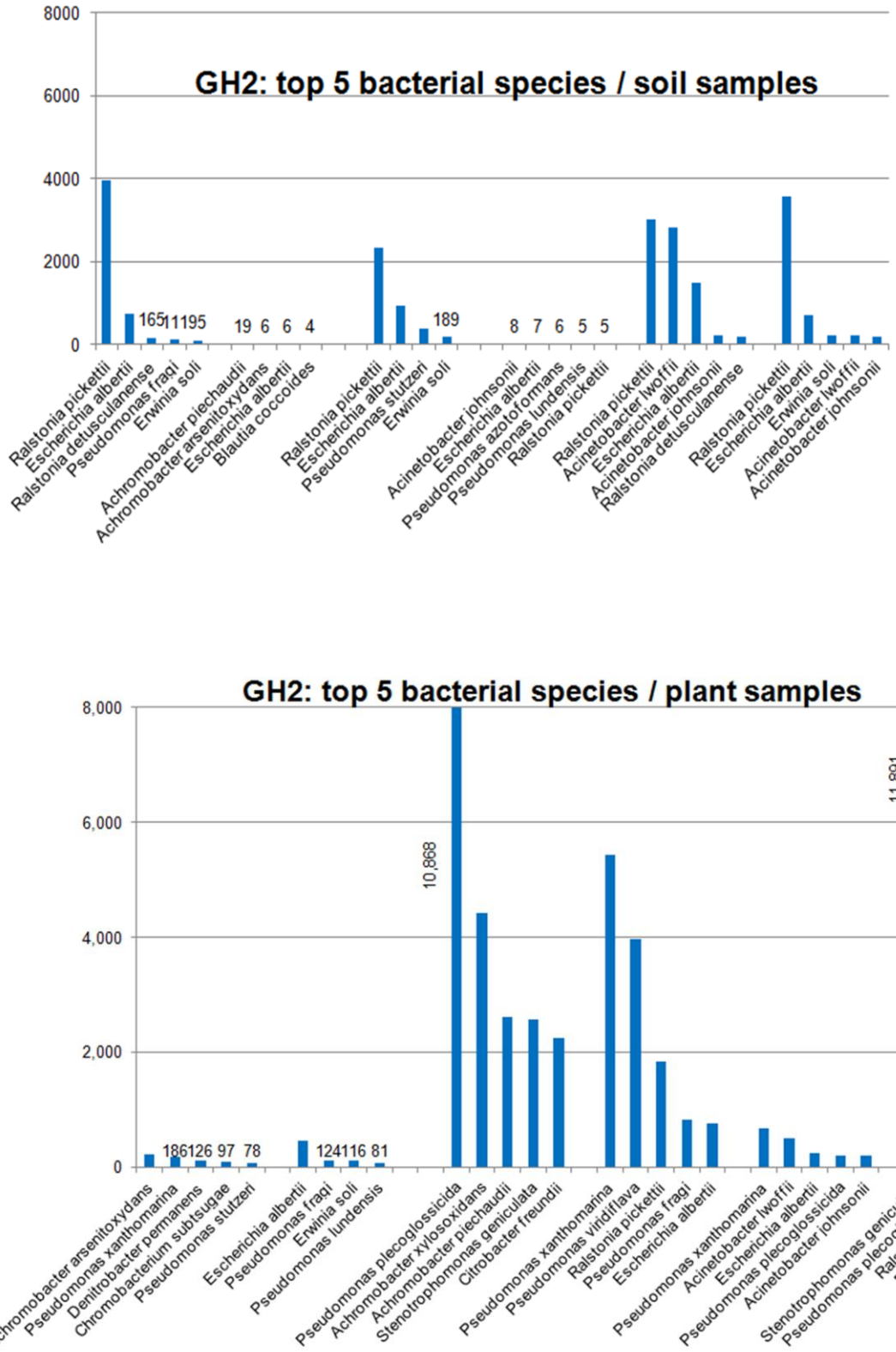


Figure 10: The most abundant five bacterial species obtained in each visit in soil and plant leaves in green house 2 (GH2).

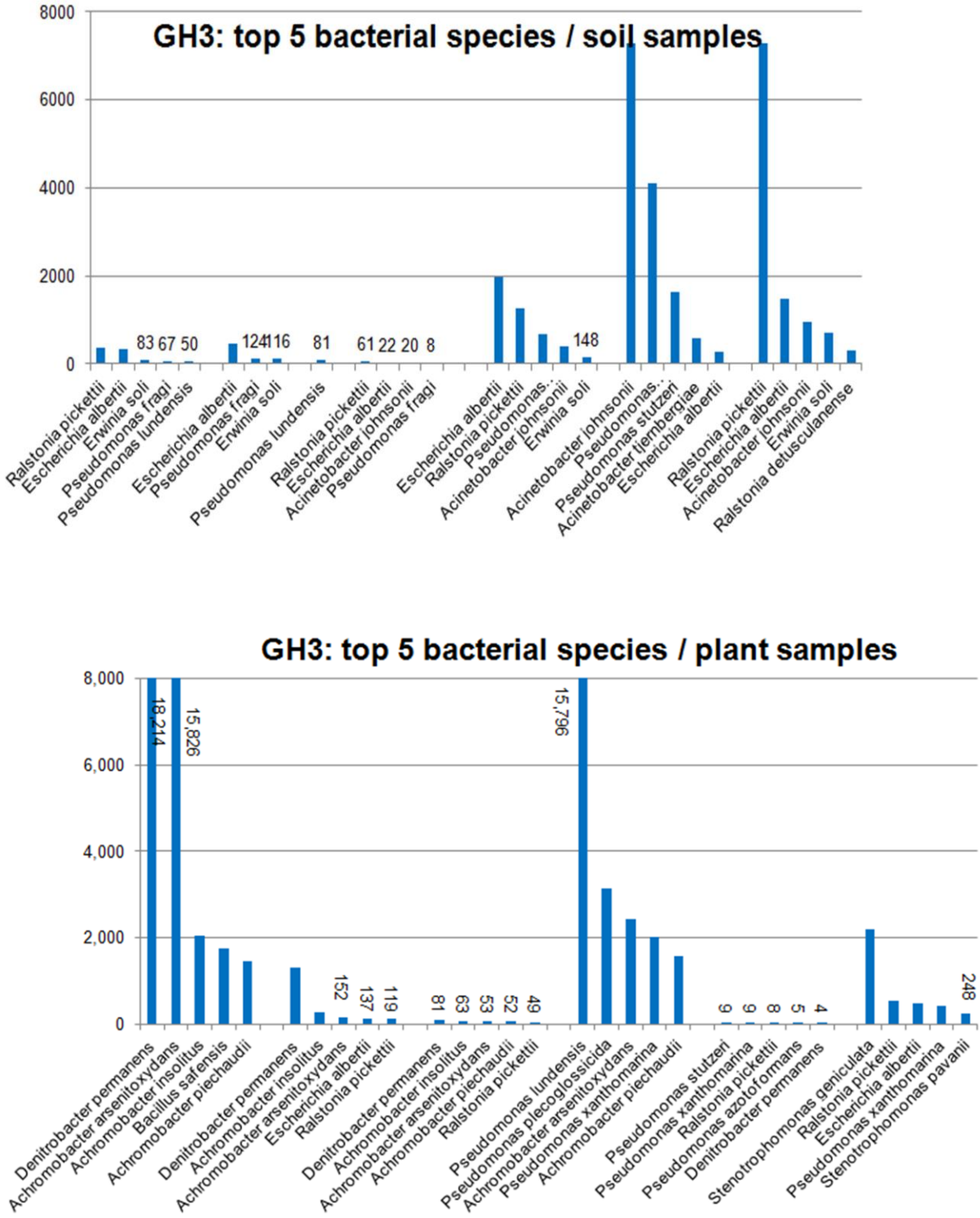


Figure 11: The most abundant five bacterial species obtained in each visit in soil and plant leaves in green house 3 (GH3).

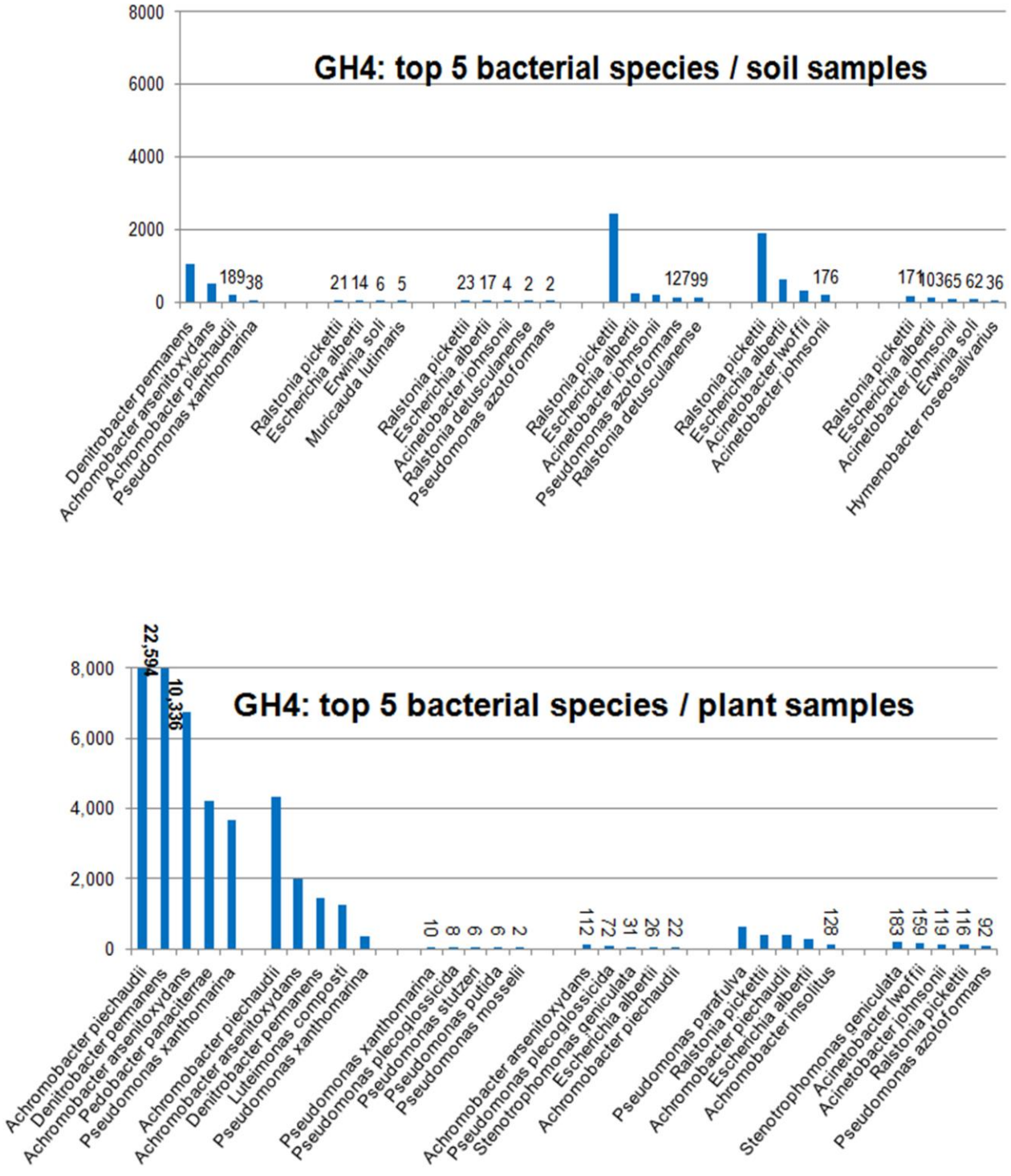


Figure 12: The most abundant five bacterial species obtained in each visit in soil and plant leaves in green house 4 (GH4).

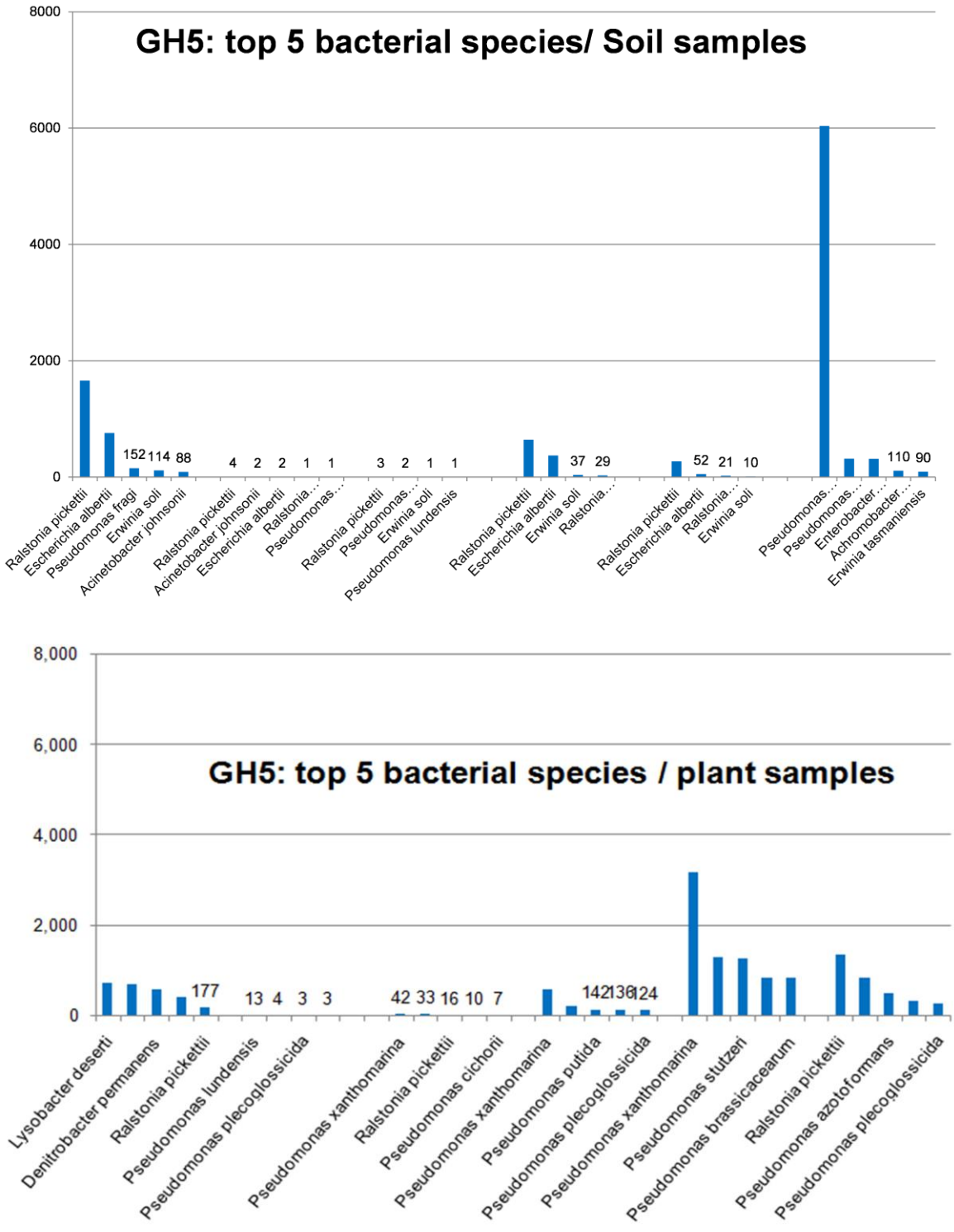


Figure 13: The most abundant five bacterial species obtained in each visit in soil and plant leaves in green house 5 (GH5).

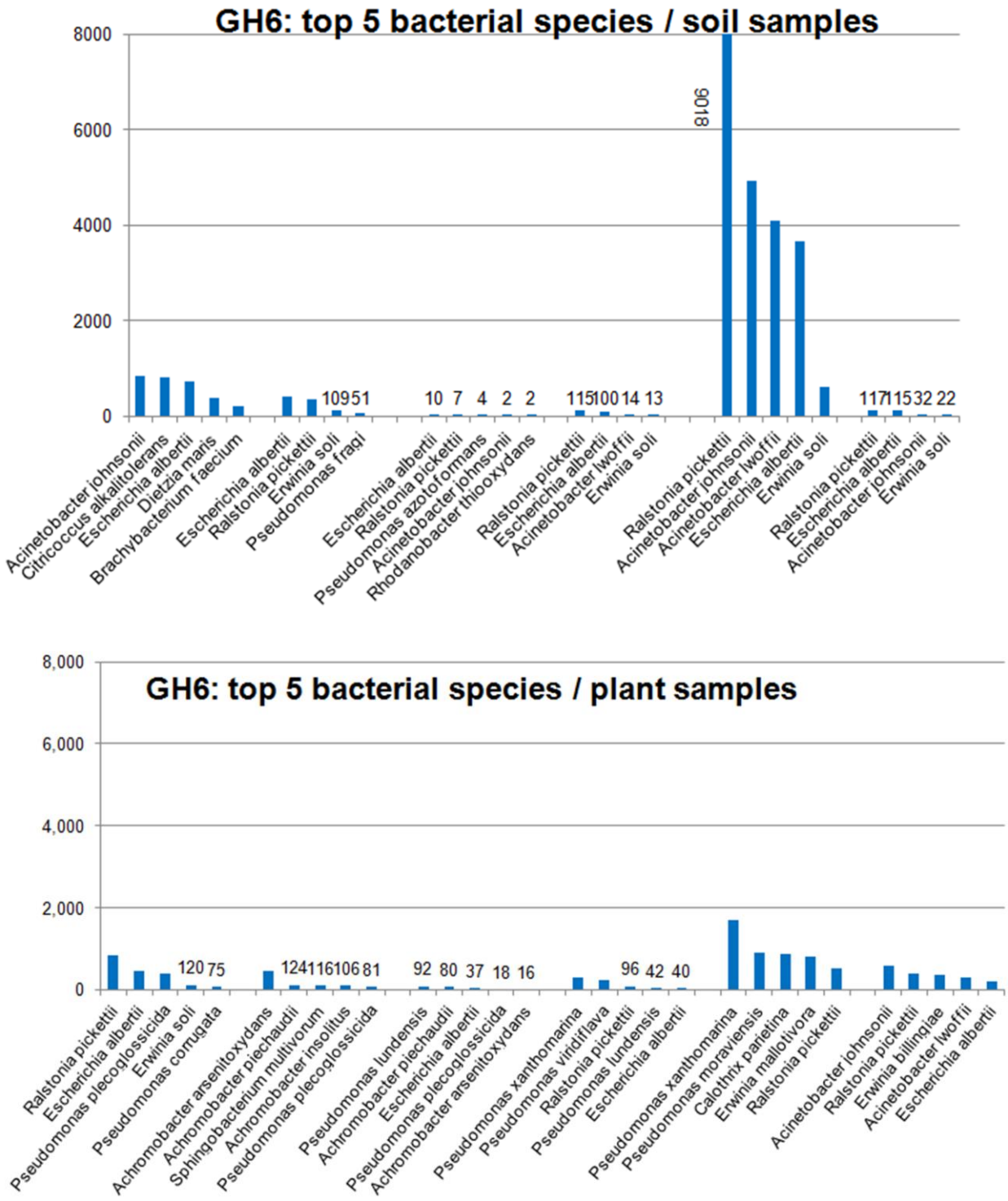


Figure 14: The most abundant five bacterial species obtained in each visit in soil and plant leaves in green house 6 (GH6).

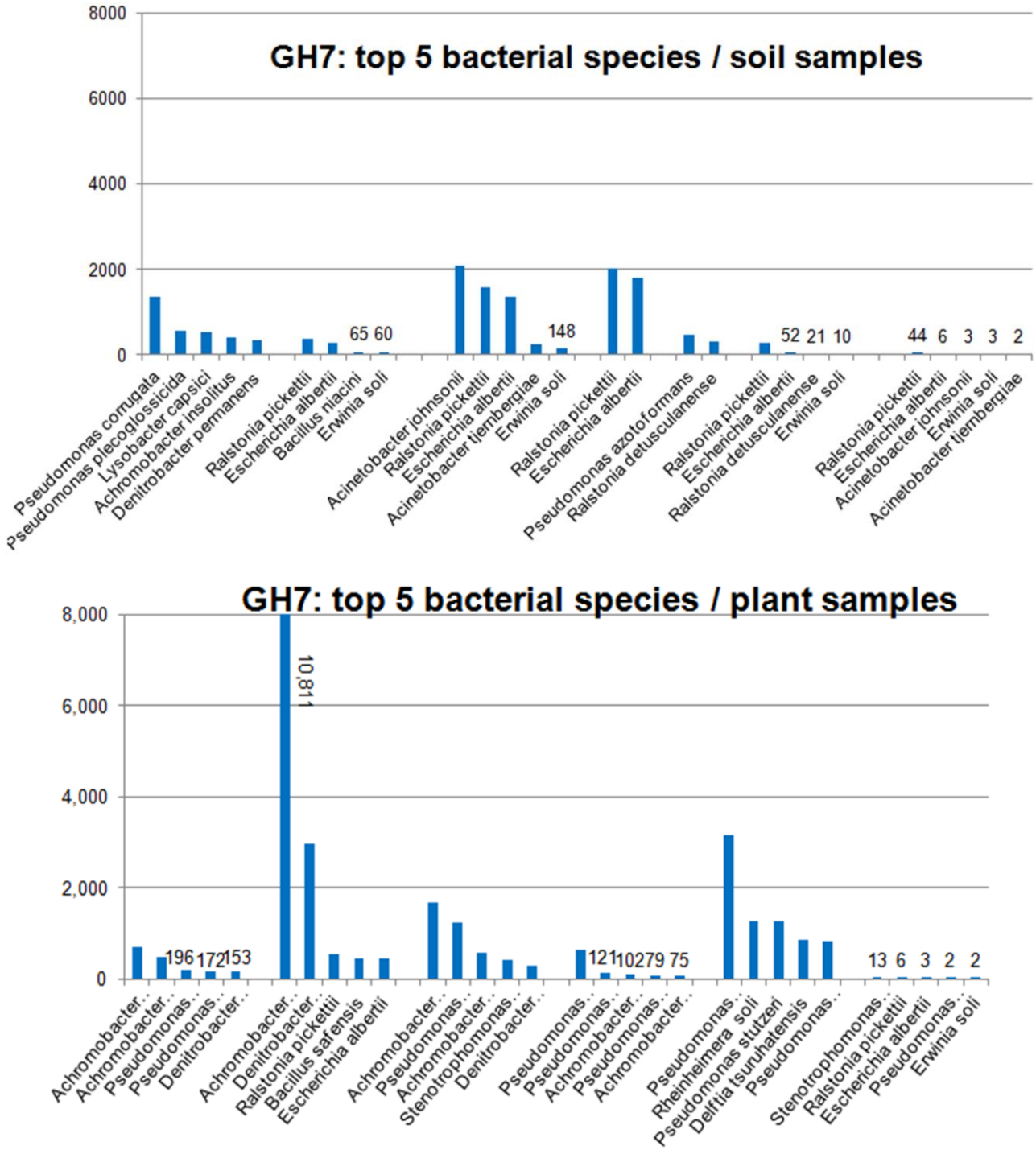


Figure 15: The most abundant five bacterial species obtained in each visit in soil and plant leaves in green house 7 (GH7).

2.3. NGS Fungi analysis in soil and plant samples: The total number of reads that represents fungi were found to be low, and many samples showed a negative finding regarding fungi species. Figure 16 shows the results of total number of reads that were obtained per each green-house in soil or plant leaves samples. The results shows consistent pattern between fungi in soil and plant leaves as well. This pattern was most prominent in visit 5 and in green-house 1 and 5, so whenever there was an increase in fungi reads in soil samples this was also reflected as a similar increase in plant leaves samples.

The main plant pathogen fungi that was identified in both soil and plant leaves was *Alternaria tenuissima*. Other fungi species were also found in one of the samples or both of them (Table 3).

Table3: List of indentified fungi in soil or plant leaves samples.

Fungi found in soil samples	Fungi found in leaves samples
Candida sake	*Alternaria tenuissima*
Geotrichum candidum	*Candida sake*
Galactomyces pseudocandidum	*Yarrowialipolytica*
Metschnikowia sp	*Galactomyces pseudocandidum*
Candida / pseudolambica	*Wickerhamiella pararugosa*
Ebaryomyces hansenii	*Candida / pararugosa*
Candida pararugosa	

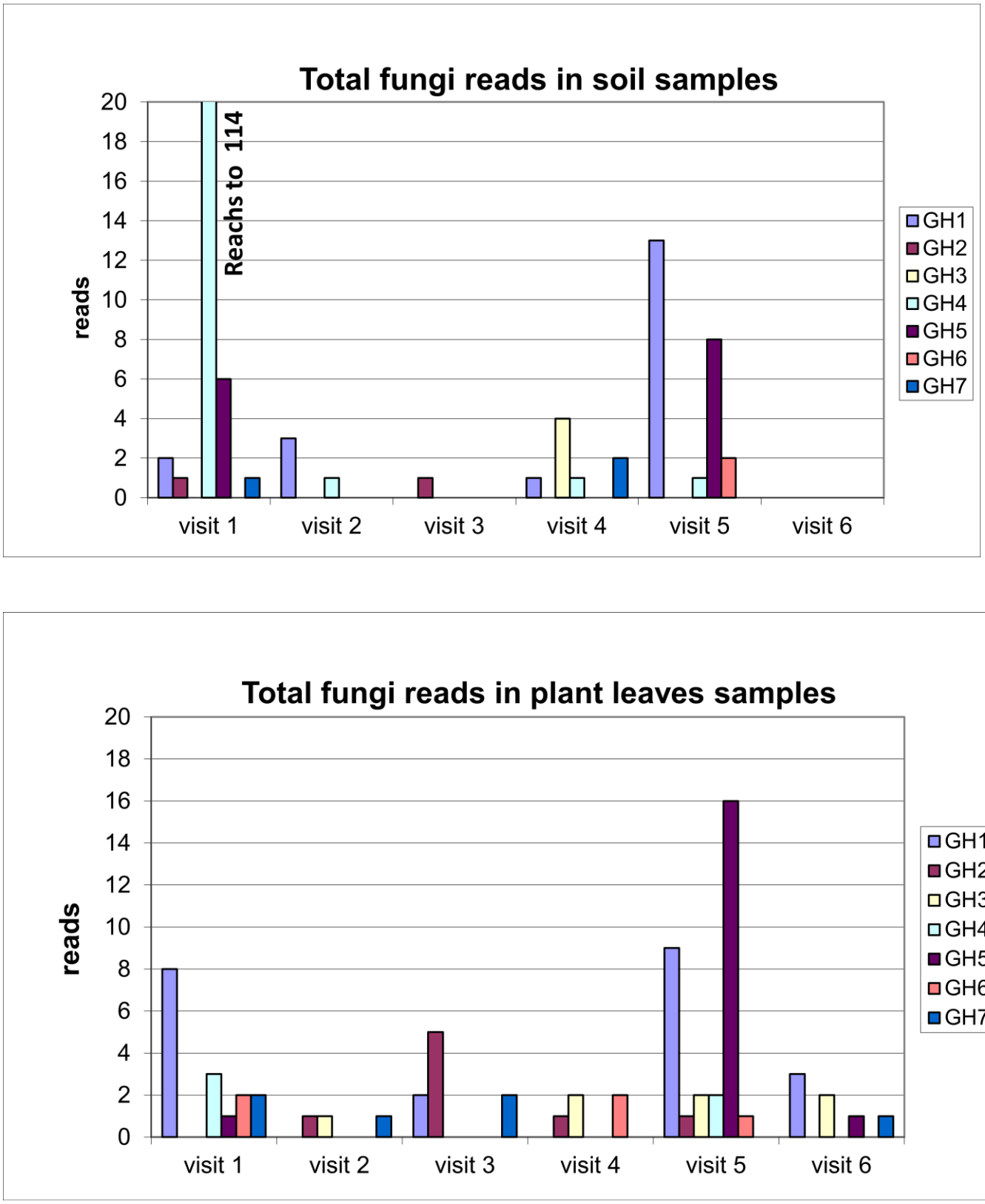


Figure 16: Total number of fungi reads in different green houses isolated from soil and plant leaves samples.

Chapter four: Discussion

Agribusiness is one of the most important arms of the Palestinian national economy that could reduce unemployment among youth and many university graduates with related speciality. In order to have a sustainable agricultural crop production there is a need for the use of crops and varieties that have a sustainable production of healthy plants and produces a higher nutritional value and productive plants. The large scale use of chemical pesticides against many types of insects and plant pests valued an annual millions of dollars and cause a serious environmental hazard, and results in development of resistance in pests against these widely used chemicals. The early detection of different plant pathogenic diseases can significantly reduce the amount of used pesticides and bringing a safer agricultural product.

Next Generation Sequencing (NGS)) or called high throughput sequencing method is relatively new technology that allows mass sequencing of different DNA fragments from an amplification reaction of fragmentation of whole genome (Besser et al., 2018; Bonk et al., 2018; Mardis, 2013). This technology enables the production of a vast array of genomic information from many organisms in parallel and it is provides a separate quantitative counting measurement for each sequenced DNA segment type (Bonk et al., 2018; Klindworth et al., 2013; Salipante, 2013). Adapting NGS technology reduces the cost of DNA sequencing compared to Sanger traditional method taking into account the large number of samples that can be analysed at once and avoiding time-consuming and tedious traditional cloning steps for sequencing of each single amplified DNA fragment. The main advantage of NGS technology is its high sensitivity and sequencing accuracy (de Paz et al., 2018; Zhu et al., 2017) since it can effectively provide a huge information on short sequence repeats or single nucleotide polymorphism (SNP) as it is provides sequence information for many DNA molecules of the same type. Using NGS sequencing approach

it is possible to use universal primers that amplify a group of pathogens such as the 16S rRNA genes for bacterial species (Salipante, 2013), the 18S rRNA for fungal pathogens (Colabella et al., 2018; Imabayashi et al., 2016), the universal primers will amplify all related DNA segments, and later all produced amplicons could be sequenced independently from each other. The obtained results will include thousand of sequences from each type of amplicon that reflects its abundance and nature.

Next-generation sequencing (NGS) has enabled in-depth investigations of the microbial communities associated with animals, plants, and fungi. Another important advantage of NGS applications is in plant breeders and the discovery of genetic variation. It is possible to sequence rapidly multiple plant species and varieties at minimal cost, and then performing the required bioinformatics analysis in order to find new sequences (simple sequence repeats (SSRs) and SNPs) that have advantageous criteria over other varieties in terms of pathogen resistance and salt tolerance (Perez-de-Castro et al., 2012). The major added value of NGS analysis over its sensitivity is of its being a quantitative method. The importance of this criteria was clearly demonstrated in identification of soil and plant microbiome and fungal inhabitant species. The advantage of using NGS technology is related to the method capability to sequence each single amplicon, from one side it can identify any mixed pathogenic organisms such as mixed vabterial infections and this by sequencing the a unique 16s rDNA gene and from the other side it will give the number of the amplified amplicons.

In this study we concentrated on monitoring the microbial and fungal infections of tomato (*Solanumlyco persicum*) crops since it considered as one of the most important vegetables, with a global annual yield of over 100 million tons (FAO. Statistical database

[FAOSTAT], 2014). Tomato production is mainly affected by bacterial infection caused by *Ralstonia solanacearum* and *Pseudomonas* causing a disease known as tomato wilt (Deberdt, 1999) This main pathogen affecting tomato is a soil-borne pathogen that also responsible on infections caused in many other plant species, *e.g.*, olive, tomato, tobacco, and eggplant, and causing a significant economical loss (Kelman, 1998). Principally this disease can be managed through cultivation in pathogen free soil using classical methods of soil fumigation, soil amendment, crop rotation, and field sanitation (Yuliar et al., 2015).

The main objectives of microbial and fungal tomato crop monitoring is to find if there is any association between soil microbiome and fungal pathogens and that microbiome found in tomato plant parts. The study supports the transfer of microbiome species from soil to plant green parts, as it was recorded figure 7 that showed the microbial abundance in plant leaves is higher in soil samples in most of the studied green-houses (GH1, GH2, and GH3). While green-house 4-7 showed consistent association of microbiome abundance in both soil and plant leaves (Figure 7). The fact that GH3 and GH4 have higher soil microbiome at the beginning of the cultivation period is later reflected in more micriobiome species in plant leaves (Figure 6-8), this supports a soil bacterial contamination that later affecting the vegitative plants. It is worth to indicate that in all studied green-houses the farmers used untreated manure compost as natural fertilizers added to soil that considered a main factor of exogenic bacterial source of contamination. Also another important factor in some occasions farmers are using waste water for plant irrigation (farmer personal communications), that also adding an exogenic pathogen source. Another important evidence that supports the soil as being a main source of bacterial infection is what was observed in GH4 and GH5 where there was a low microbiome abundance in both soil and plant leaves samples, which showed a healthy growth all over the study period (Figure 8). The assumption of soil being a source of infection also was

discussed by other researchers (Lee CG, 2017). Also, other studies compared the microbiome and fungal communities in soil and plant samples looking for some bio-control microbial or fungal agents that compete with the pathogenic micro-organisms (Chialva M, 2018). Lee *et al*, 2017; concluded that prokaryotic and eukaryotic community structures were affected by regional differences than the appearance of disease, and they identified several prokaryotes and eukaryotes that were more abundant in soil that lacked disease symptoms. And in another study they found that tomato plants grown on native soils with their complex microbes respond differently from tomato growing in a sterile substrate(Chialva M, 2018). They showed that soils that have *Fusarium oxysporum* are triggering immunity pathway as the first level of plant defence and causing lignin synthesis leading to active protection to growing tomato plants.

The specific microbiome species were of most imoprtance in this study, several plant pathogenic bacterial species were idenitfed, of most important species are *Ralstonia, Erwina, Pseudomonas, Stenotrophomonas,* and *Achromobacter* (Figures 9-14). These species are of soil origin and causing different diseases in tomato plant, most important is *Ralstonia* bacterial species that cause tomato wilt disease. The fact that it is not always we find the main pathogenic *Ralstonia solanacearum*, but we find other species is not eliminating of having this pathogenic species since there is 99% similarity in 16s rDNA among most *Ralstonia* bacterial species.

As indicating before many of the found bacterial species are important as plant pathogens and there is a need for more evaluation of their abundance and their major effect on plant pathogenesis. The following is a brief description for the most important found bacterial species.

- ***Erwinia*** is a genus of Entero bacteriaceae Gram-negative rod-shaped bacteria containing many plant pathogenic species.

- ***Achromobacterar senitoxydans:***A soil Gram-negative, rod-shaped bacterium normaly isolated from soil that contaminated with animal minure.

- ***Pseudomonas xanthomarina*** is a Gram-negative, rod-shaped, Isolated from animals, plants, and polluted soils.

- ***Stenotrophomona sgeniculata*** is a of Gram-negative bacteria normaly considered a common soil organisms and it could be opportunistic human pathogens.

- ***Achromobacter xylosoxidans*** is Gram-negative bacillus that colonizes aquatic environments. This bacteria has been described as an opportunistic human pathogen.

- ***Pseudomonas fragi:*** Gram-negative bacterium commonly found in soil.

- ***Achromobacter****is* Gram-negative straight rods found in water and soils and are opportunistic human pathogens.

References

Abbasi, I., Aramin, S., Hailu, A., Shiferaw, W., Kassahun, A., Belay, S., Jaffe, C., Warburg, A., 2013. Evaluation of PCR procedures for detecting and quantifying Leishmania donovani DNA in large numbers of dried human blood samples from a visceral leishmaniasis focus in Northern Ethiopia. BMC infectious diseases 13, 153.

Abbasi, I., Trancoso Lopo de Queiroz, A., Kirstein, O.D., Nasereddin, A., Horwitz, B.Z., Hailu, A., Salah, I., Mota, T.F., Fraga, D.B.M., Veras, P.S.T., Poche, D., Poche, R., Yeszhanov, A., Brodskyn, C., Torres-Poche, Z., Warburg, A., 2018. Plant-feeding phlebotomine sand flies, vectors of leishmaniasis, prefer Cannabis sativa. Proceedings of the National Academy of Sciences of the United States of America 115, 11790-11795.

Abd-Elsalam, K.A., 2003. Non-gel based techniques for plant pathogen genotyping. Acta microbiologica Polonica 52, 329-341.

Abed, A., Legare, G., Pomerleau, S., St-Cyr, J., Boyle, B., Belzile, F.J., 2019. Genotyping-by-Sequencing on the Ion Torrent Platform in Barley. Methods in molecular biology 1900, 233-252.

Aly, S.M., Sabri, D.M., 2015. Next generation sequencing (NGS): a golden tool in forensic toolkit. Archiwum medycyny sadowej i kryminologii 65, 260-271.

Ambardar, S., Gupta, R., Trakroo, D., Lal, R., Vakhlu, J., 2016. High Throughput Sequencing: An Overview of Sequencing Chemistry. Indian journal of microbiology 56, 394-404.

Ameur, A., Kloosterman, W.P., Hestand, M.S., 2018. Single-Molecule Sequencing: Towards Clinical Applications. Trends in biotechnology.

Ayliffe, M., Periyannan, S.K., Feechan, A., Dry, I., Schumann, U., Lagudah, E., Pryor, A., 2014. Simple quantification of in planta fungal biomass. Methods in molecular biology 1127, 159-172.

Barnes, R.A., White, P.L., 2016. PCR Technology for Detection of Invasive Aspergillosis. Journal of fungi 2.

Bartlett, J.M.S., Stirling, David 2003. PCR Protocols. Springer.
Begerow, D., Nilsson, H., Unterseher, M., Maier, W., 2010. Current state and perspectives of fungal DNA barcoding and rapid identification procedures. Applied microbiology and biotechnology 87, 99-108.

Berg, G., Rybakova, D., Grube, M., Koberl, M., 2016. The plant microbiome explored: implications for experimental botany. Journal of experimental botany 67, 995-1002.

Besser, J., Carleton, H.A., Gerner-Smidt, P., Lindsey, R.L., Trees, E., 2018. Next-generation sequencing technologies and their application to the study and control of bacterial infections. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases 24, 335-341.

Bonk, F., Popp, D., Harms, H., Centler, F., 2018. PCR-based quantification of taxa-specific abundances in microbial communities: Quantifying and avoiding common pitfalls. Journal of microbiological methods 153, 139-147.

Bulgarelli D, S.K., Spaepen S. Ver Loren van Themaat E, Schulze-LefertP. , 2013. Structure and functions of the bacterial microbiota of plants. Annu Rev Plant Biol 64, 807–838.

Burks, C., Cassidy, M., Cinkosky, M.J., Cumella, K.E., Gilna, P., Hayden, J.E., Keen, G.M., Kelley, T.A., Kelly, M., Kristofferson, D., et al., 1991. GenBank. Nucleic acids research 19 Suppl, 2221-2225.

Burks, C., Fickett, J.W., Goad, W.B., Kanehisa, M., Lewitter, F.I., Rindone, W.P., Swindell, C.D., Tung, C.S., Bilofsky, H.S., 1985. The GenBank nucleic acid sequence database. Computer applications in the biosciences : CABIOS 1, 225-233.

Burks, C., Tomlinson, L.J., 1989. Submission of data to GenBank. Proceedings of the National Academy of Sciences of the United States of America 86, 408.

Chen, G., Qiu, Y., Zhuang, Q., Wang, S., Wang, T., Chen, J., Wang, K., 2018. Next-generation sequencing library preparation method for identification of RNA viruses on the Ion Torrent Sequencing Platform. Virus genes.

Chialva M, Z.Y., Spadaro D, Bonfante P., 2018. Not only priming: Soil microbiota may protect tomato from root pathogens. Plant Signal Behav 13, e1464855.

Colabella, C., Corte, L., Roscini, L., Bassetti, M., Tascini, C., Mellor, J.C., Meyer, W., Robert, V., Vu, D., Cardinali, G., 2018. NGS barcode sequencing in taxonomy and diagnostics, an application in "Candida" pathogenic yeasts with a metagenomic perspective. IMA fungus 9, 91-105.

Cuartero, J., Bolarin, M.C., Asins, M.J., Moreno, V., 2006. Increasing salt tolerance in the tomato. Journal of experimental botany 57, 1045-1058.

Davidson, R.M., Epperson, L.E., 2018. Microbiome Sequencing Methods for Studying Human Diseases. Methods in molecular biology 1706, 77-90.

de Paz, A.M., Cybulski, T.R., Marblestone, A.H., Zamft, B.M., Church, G.M., Boyden, E.S., Kording, K.P., Tyo, K.E.J., 2018. High-resolution mapping of DNA polymerase fidelity using nucleotide imbalances and next-generation sequencing. Nucleic acids research 46, e78.

Deberdt, P., P. Quénéhervé, A. Darrasse, and P. Prior., 1999. ncreased susceptibility to bacterial wilt in tomatoes by nematode galling and the role of the Mi gene in resistance to nematodes and bacterial wilt. Plant Pathol. 48, 408-414.

Elnesr, M.N., Alazba, A.A., Zein El-Abedein, A.I., El-Adl, M.M., 2015. Evaluating the Effect of Three Water Management Techniques on Tomato Crop. PloS one 10, e0129796.

Gabriel, C., Furst, D., Fae, I., Wenda, S., Zollikofer, C., Mytilineos, J., Fischer, G.F., 2014. HLA typing by next-generation sequencing - getting closer to reality. Tissue antigens 83, 65-75.

Ikeda-Ohtsubo, W., Brugman, S., Warden, C.H., Rebel, J.M.J., Folkerts, G., Pieterse, C.M.J., 2018. How Can We Define "Optimal Microbiota?": A Comparative Review of Structure and Functions of Microbiota of Animals, Fish, and Plants in Agriculture. Frontiers in nutrition 5, 90.

Imabayashi, Y., Moriyama, M., Takeshita, T., Ieda, S., Hayashida, J.N., Tanaka, A., Maehara, T., Furukawa, S., Ohta, M., Kubota, K., Yamauchi, M., Ishiguro, N., Yamashita, Y., Nakamura, S., 2016. Molecular analysis of fungal populations in patients with oral candidiasis using next-generation sequencing. Sci Rep 6, 28110.

Jo, J.H., Kennedy, E.A., Kong, H.H., 2016. Research Techniques Made Simple: Bacterial 16S Ribosomal RNA Gene Sequencing in Cutaneous Research. The Journal of investigative dermatology 136, e23-e27.

Kelman, A., 1998. One hundred and one years of research on bacterial wilt, in: P.H. Prior, C.A., and J. Elphinstone (Ed.), Bacterial Wilt Disease: Molecular and Ecological Aspects. Springer, Heidelberg.

Kemen, A.C., Agler, M.T., Kemen, E., 2015. Host-microbe and microbe-microbe interactions in the evolution of obligate plant parasitism. The New phytologist 206, 1207-1228.

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, e1.

Knief, C., 2014. Analysis of plant microbe interactions in the era of next generation sequencing technologies. Frontiers in plant science 5, 216.

Kuzdralinski, A., Kot, A., Szczerba, H., Nowak, M., Muszynska, M., 2017. A Review of Conventional PCR Assays for the Detection of Selected Phytopathogens of Wheat. Journal of molecular microbiology and biotechnology 27, 175-189.

Ledergerber, C., Dessimoz, C., 2011. Base-calling for next-generation sequencing platforms. Briefings in bioinformatics 12, 489-497.

Lee CG, I.T., Uwagaki Y, Otani Y, Nakaho K, Ohkuma M., 2017. Comparison of Prokaryotic and Eukaryotic Communities in Soil Samples with and without Tomato Bacterial Wilt Collected from Different Fields. Microbes and environments Dec;32, 376-385.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., Law, M., 2012. Comparison of next-generation sequencing systems. Journal of biomedicine & biotechnology 2012, 251364.

López MM, B.E., Olmos A, Caruso P, Gorris MT, Llop P, Penyalver R, Cambra M., 2003. Innovative tools for detection of plant pathogenic viruses and bacteria. Int Microbiol. 6, 233-243.

Mardis, E.R., 2013. Next-generation sequencing platforms. Annual review of analytical chemistry (Palo Alto, Calif.) 6, 287-303.

Massart, S., Olmos, A., Jijakli, H., Candresse, T., 2014. Current impact and future directions of high throughput sequencing in plant virus diagnostics. Virus research 188, 90-96.

Maxwell, P., Hynes, S.O., Fuchs, M., Craig, S., McGready, C., McLean, F., McQuaid, S., James, J., Salto-Tellez, M., 2018. Practical guide for the comparison of two next-generation sequencing systems for solid tumour analysis in a universal healthcare system. Journal of clinical pathology.

McCartney, H.A., Foster, S.J., Fraaije, B.A., Ward, E., 2003. Molecular diagnostics for fungal plant pathogens. Pest management science 59, 129-142.

Mendes, R., Kruijt, M., de Bruijn, I., Dekkers, E., van der Voort, M., Schneider, J.H., Piceno, Y.M., DeSantis, T.Z., Andersen, G.L., Bakker, P.A., Raaijmakers, J.M., 2011. Deciphering the rhizosphere microbiome for disease-suppressive bacteria. Science 332, 1097-1100.

Merriman, B., Ion Torrent, R., Team, D., Rothberg, J.M., 2012. Progress in ion torrent semiconductor chip based sequencing. Electrophoresis 33, 3397-3417.

Nezhad, A.S., 2014. Future of portable devices for plant pathogen diagnosis. Lab on a chip 14, 2887-2904.

Nolasco, G., de Blas, C., Torres, V., Ponz, F., 1993. A method combining immunocapture and PCR amplification in a microtiter plate for the detection of plant viruses and subviral pathogens. Journal of virological methods 45, 201-218.

O'Sullivan, C.E., Kasai, M., Francesconi, A., Petraitis, V., Petraitiene, R., Kelaher, A.M., Sarafandi, A.A., Walsh, T.J., 2003. Development and validation of a quantitative real-time PCR assay using fluorescence resonance energy transfer technology for detection of Aspergillus fumigatus in experimental invasive pulmonary aspergillosis. Journal of clinical microbiology 41, 5676-5682.

Perez-de-Castro, A.M., Vilanova, S., Canizares, J., Pascual, L., Blanca, J.M., Diez, M.J., Prohens, J., Pico, B., 2012. Application of genomic tools in plant breeding. Current genomics 13, 179-195.

Pravi, V., Jeeva, M.L., Archana, P.V., 2015. Nucleic acid spot hybridization based species-specific detection of Sclerotium rolfsii associated with collar rot disease of Amorphophallus paeoniifolius. World journal of microbiology & biotechnology 31, 315-320.

Ravi, R.K., Walton, K., Khosroheidari, M., 2018. MiSeq: A Next Generation Sequencing Platform for Genomic Analysis. Methods in molecular biology 1706, 223-232.

Ray M, R.A., Dash S, Mishra A, Achary KG, Nayak S, Singh S., 2017. Fungal disease detection in plants: Traditional assays, novel diagnostic techniques and biosensors. Biosens Bioelectron 15, 708-723.

Rettcher, S., Jungk, F., Kuhn, C., Krause, H.J., Nolke, G., Commandeur, U., Fischer, R., Schillberg, S., Schroper, F., 2015. Simple and portable magnetic immunoassay for rapid detection and sensitive quantification of plant viruses. Applied and environmental microbiology 81, 3039-3048.

Roossinck, M.J., 2017. Deep sequencing for discovery and evolutionary analysis of plant viruses. Virus research 239, 82-86.

Ropokis, A., Ntatsi, G., Kittas, C., Katsoulas, N., Savvas, D., 2018. Impact of Cultivar and Grafting on Nutrient and Water Uptake by Sweet Pepper (Capsicum annuum L.) Grown Hydroponically Under Mediterranean Climatic Conditions. Frontiers in plant science 9, 1244.

Salipante, S.J., Sengupta, D.J., Rosenthal, C., Costa, G., Spangler, J., Sims, E.H., Jacobs, M.A., Miller, S.I., Hoogestraat, D.R., Cookson, B.T., McCoy, C., Matsen, F.A., Shendure, J., Lee, C.C., Harkins, T.T., Hoffman, N.G, 2013. Rapid 16S rRNA next-generation sequencing of polymicrobial clinical samples for diagnosis of complex bacterial infections. PLoS One 8, e65226.

Schaad, N., Schuenzel, E., 2010. Sensitive molecular diagnostic assays to mitigate the risks of asymptomatic bacterial diseases of plants. Critical reviews in immunology 30, 271-275.
Sharma, K.K., 2016. Fungal genome sequencing: basic biology to biotechnology. Critical reviews in biotechnology 36, 743-759.

Tedersoo, L., Drenkhan, R., Anslan, S., Morales-Rodriguez, C., Cleary, M., 2018. High-throughput identification and diagnostics of pathogens and pests: Overview and practical recommendations. Molecular ecology resources.

Thornton, C.R., 2004. An immunological approach to quantifying the saprotrophic growth dynamics of Trichoderma species during antagonistic interactions with Rhizoctonia solani in a soil-less mix. Environmental microbiology 6, 323-334.

Todd, J.R., Matsumoto, T., Ueno, R., Murugaiyan, J., Britten, A., King, J.W., Odaka, Y., Oberle, A., Weise, C., Roesler, U., Pore, R.S., 2018. Medical phycology 2017. Medical mycology 56, S188-S204.

Tsui, C.K., Woodhall, J., Chen, W., Levesque, C.A., Lau, A., Schoen, C.D., Baschien, C., Najafzadeh, M.J., de Hoog, G.S., 2011. Molecular techniques for pathogen identification and fungus detection in the environment. IMA fungus 2, 177-189.

Walker, S.E., Lorsch, J., 2013. Sanger dideoxy sequencing of DNA. Methods in enzymology 529, 171-184.

Xu, J., 2016. Fungal DNA barcoding. Genome 59, 913-932.

Yuliar, Nion, Y.A., Toyota, K., 2015. Recent trends in control methods for bacterial wilt diseases caused by Ralstonia solanacearum. Microbes and environments 30, 1-11.

Zhu, W., Zhang, X.Y., Marjani, S.L., Zhang, J., Zhang, W., Wu, S., Pan, X., 2017. Next-generation molecular diagnosis: single-cell sequencing from bench to bedside. Cellular and molecular life sciences : CMLS 74, 869-880.