



Metagenomics study of endophytic bacteria in *Aloe vera* using next-generation technology



Mushafau Adewale Akinsanya^{a,b}, Joo Kheng Goh^a, Siew Ping Lim^a, Adeline Su Yien Ting^{a,*}

^a School of Science, Monash University Malaysia, 46150 Bandar Sunway, Selangor, Malaysia

^b Department of Medical Biochemistry, Faculty of Basic Medical Sciences, College of Medicine, Lagos State University, P.M.B 21266 Ikeja, Lagos, Nigeria

ARTICLE INFO

Article history:

Received 18 July 2015

Received in revised form 19 August 2015

Accepted 4 September 2015

Available online 10 September 2015

Keywords:

Aloe vera

α -Diversity

Bacterial endophytes

Illumina

Metagenomics

NGS

ABSTRACT

Next generation sequencing (NGS) enables rapid analysis of the composition and diversity of microbial communities in several habitats. We applied the high throughput techniques of NGS to the metagenomics study of endophytic bacteria in *Aloe vera* plant, by assessing its PCR amplicon of 16S rDNA sequences (V3–V4 regions) with the Illumina metagenomics technique used to generate a total of 5,199,102 reads from the samples. The analyses revealed *Proteobacteria*, *Firmicutes*, *Actinobacteria* and *Bacteroidetes* as the predominant genera. The roots have the largest composition with 23% not present in other tissues. The stems have more of the genus—*Pseudomonas* and the unclassified *Pseudomonadaceae*. The α -diversity analysis indicated the richness and inverse Simpson diversity index of the bacterial endophyte communities for the leaf, root and stem tissues to be 2.221, 6.603 and 1.491 respectively. In a similar study on culturable endophytic bacteria in the same *A. vera* plants (unpublished work), the dominance of *Pseudomonas* and *Bacillus* genera was similar, with equal proportion of four species each in root, stem and leaf tissues. It is evident that NGS technology captured effectively the metagenomics of microbiota in plant tissues and this can improve our understanding of the microbial–plant host interactions.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The diversity of microorganisms on earth remains poorly understood although an estimated 1.5 million species of bacteria and fungi have vital functions as decomposers, symbionts, and pathogens in ecosystems [1]. To date, only 5% of the estimated number of bacterial species has been documented. In recent years, metagenomics studies have improved our understanding of the diversity of microbes in various habitats. This includes microbes associated with plants, which thrive below ground in the rhizosphere, above in the phyllosphere [2] and within the plant tissues as endophytes [3,4]. These microbes can have beneficial, neutral, or detrimental effects on the plant. The association between microbiota and plants is important, as it leads to understanding microbes and “What are they doing and how do they respond to environmental changes and interact with each other?” These could further contribute to understanding the significant roles of plant microbiota in supporting plant growth and improved crop yield.

Genomic analyses of individual strains or metagenomics studies of whole microbial communities may provide insight into the composition or diversity and physiological potential of endophytes associated with plants. For example, the study of microbial (endophyte) diversity in

plants tissues, reveals both culturable and unculturable endophytes that may be beneficial microbes, subsequently gearing towards their isolation and characterisation. It is also possible to further evaluate evolutionary trend of the associated microbes and how they are related with one another. We may also be able to evaluate the statement of their close endophyte–host association and co-evolution in relation to their ability to produce similar compounds to that of their host [3].

In recent studies, endophytes have been shown to have an important role in promoting plant growth and yield, suppress pathogens, aid in removing contaminants, solubilize phosphate or contribute to nitrogen assimilation for plants [5,6]. Over the past decade, our understanding of microbial diversity and function in complex environments has increased significantly, primarily as a result of the introduction of next generation sequencing (NGS) [7]. Both PCR based analysis of 16S rRNA gene and shotgun metagenomics studies have been used recently to characterise soils [8], oceans [9], the atmosphere, as well as the human microbiome [10]. Prior to the introduction of NGS, these characterisations were done at extremely high cost [11]. The use of the Illumina platform to generate data sets of unprecedented size [12,13] further revealed more information of various microbiomes but at much lower cost.

Next generation sequencing of hypervariable regions from small-subunit ribosomal RNA genes (16S rRNA) is useful for analyses of microbial communities in several habitats [14]. The use of high-throughput short-read sequencing of the 16S rRNA amplicon for the profiling of

* Corresponding author.

E-mail address: adeline.ting@monash.edu (A.S.Y. Ting).

microbial communities has become an increasingly attractive option by researchers as the amplicon consists of the conserved region interspersed by variable regions that facilitate sequencing and phylogenetic classification. The 454 GS-20 pyrosequencing by Roche in 2006 was the first high-throughput sequencing technology successfully applied for biodiversity analysis, and further improvement of the technology offers read lengths of up to 1000 bp [15]. The Illumina technology is highly effective in performing comparatively high sequencing depth despite having short read lengths and reduced per base costs [9,13,16]. Therefore, this technology has been used for amplicon sequencing of bacterial and fungal marker genes to characterise microbial communities in the phyllosphere and rhizosphere [17]. In this study, Illumina technology was applied for the 16S rRNA sequencing targeting the V3–V4 regions (amplicon of 150–400 bp) using primers designed against the surrounding conserved regions [18]. The bioinformatics tools provided by mothur pipeline were explored to process raw data reads and to analyse the microbiota communities. Previous work done using Illumina platform has suggested the effectiveness of this fragment size to be sufficient for resolving microbial community differences [19].

This study focussed on endophyte communities from the *Aloe vera* plant. In our separate study (unpublished work), we have characterised some of the culturable endophyte isolates from *A. vera* with beneficial bioactive compounds. *Aloe* plants are known for its nutritional and therapeutic values. The leaf exudates are used to a great extent in traditional medicine [20]. Other uses include treating wounds and burns, also diabetes and elevated blood lipids in humans. These effects are believed to be attributed to compounds such as polysaccharides, mannans, flavonoids, anthraquinones, lectins and other phytochemical compounds that are isolated from the plant. Unmasking the overall endophytic bacteria communities may help in identifying and describing the microbial plant colonisation by both the culturable and unculturable species and their link to the bioactive compounds produced. Hence, we employed the NGS technology to unveil the culturable and unculturable endophytic bacteria in *A. vera*, and to elucidate the microbial plant colonisation pattern and evaluate its microbial diversity.

2. Results

The primary analysis of the reads through base calling directly on the MiSeq sequencing reporter (MSR), revealed raw reads statistics and sequence quality assessment as in Table 1. The fastq reads obtained per sample were in paired-end reads labelled as (L001_R1_001.fastq and L001_R2_001.fastq). The project (PRJNA288893) was registered with the GenBank, with BioSample accession numbers SAMN03839381 (root), SAMN03975610 (stem), and SAMN03975611 (leaf). The highest reads were obtained from the root tissues (2,528,030 reads) followed by the leaf (1,372,180 reads) and stem tissues (1,298,892 reads). The GC content followed the same order (52.01, 50.86 and 50.01), respectively. This trend may not be unusual since the root is closer to the soil microbial communities than other tissues. The higher reads in leaf tissues compared to the stem, may probably be the result of the relatively larger size of leaf tissues than the stem tissues, which might harbour more microbial communities.

2.1. Sequence processing

In this study, the sequences were processed using mothur, a software package with less computational demands [21]. Analysis of the

Table 2

Sequence processed details: merged sequence.

Sample reference	Before merge process Number of sequence (total sequence length in bp)	After merge process Number of sequence (total sequence length in bp)
Root	2,528,030 (361,652,861)	1,264,015 (220,836,340)
Stem	1,298,892 (191,468,046)	649,446 (124,733,765)
Leaf	1,372,180 (200,256,446)	686,090 (127,684,140)
Total		2,599,551 (473,254,245)

Sequence input (forward and reverse sequences), quality encoding (Illumina 1.8+) and Alignment method (needleman).

raw data indicated that the reads covered V3 region successfully (size ranged ~ 200 bp). Forward and reverse reads were merged and >99% were overlapped at V3 region using the mothur pipeline (Refer supplementary Figs. S1, S2 and Table 2). The merged sequences were further processed. According to Huse et al. [22], accumulation of errors within a rather small subset of 454 reads may occur hence it was necessary to remove reads with ambiguous base calls (Ns), unusual or unexpected length, low quality scores or those that cannot be aligned to the gene of interest (assumed to be unspecific PCR products) [22,23]. Reads were trimmed based on quality scores, singletons (sequence reads that occur only once) are removed from the datasets to further reduce the error rate [9].

The mothur “seqNoise algorithm” incorporated with UCHIME further removed chimeric sequences originated during PCR (5–45% of PCR product) [24,25]. UCHIME was reported to perform best in a comparative study where a reference database was used [26]. Critical analyses of different denoising tools demonstrated that parameters have to be chosen very carefully so as not to introduce bias by read modification during the generation of representative consensus reads. Hence, mothur which combined the above analyses such as OTU clustering, taxonomy assignment and multiple sample comparison, has been considered to be more appropriate or the UPARSE pipeline [26,13,27] for OTU estimation. The resulting merged sequences and processing details are shown in Tables 2 and 3 and supplementary Table S1.

2.2. Characterisation of community composition

The relative abundance of bacterial communities as obtained in the three tissues evaluated is shown in Fig. 1. Of the three tissues analysed, *Proteobacteria* sub-phylum is predominant followed by *Firmicutes*, *Actinobacteria* and *Bacteroidetes*. It was noted that the stem tissue has more of the genus—*Pseudomonas* and unclassified *Pseudomonadaceae* than the root and leaf tissues. On the contrary, leaf tissues have more of genus—*Propionibacterium*, *Serratia* and *Brevibacterium* than the root and stem tissues (refer supplementary Tables S2 and S3). In all, the root tissues have the highest richness of the four bacteria groupings.

Computational analyses of the α -diversity estimated the richness and diversity of the three samples at OTU cutoffs of 1 distance units (by using the number of observed OTUs). Chao1 estimated minimum number of OTUs, and inverse Simpson diversity index indicates the richness of the communities (refer supplementary Fig. S4). Chao1 curves continue to climb with sampling; however, the inverse Simpson diversity indices are relatively stable. The summary table of the diversity gave us the insight to the sampling coverage of the communities which is well above 99%. Also, there are significant differences of the

Table 1

Raw reads statistics and sequence quality assessment of 16S rRNA sequence from *A. vera* tissues.

Sample reference	Sample label	Sequence type	Sequence format	Read type	Read size (bp)	Total number of reads	Total sequence length (nt)	GC%
Root	22	Illumina MiSeq	Fastq	Paired-end	35–151	2,528,030	361,652,861	52.01
Stem	23					1,298,892	191,468,046	50.01
Leaf	24					1,372,180	200,256,446	50.86

Raw data from MiSeq sequencing reporter (MSR).

Download English Version:

<https://daneshyari.com/en/article/2821988>

Download Persian Version:

<https://daneshyari.com/article/2821988>

[Daneshyari.com](https://daneshyari.com)