Old Herborn University Seminar Monograph 33: The Flowering of the Plant Microbiome and the Human Connection.
Editors: Peter J. Heidt, James Versalovic, Mark S. Riddle, and Volker Rusch.
Old Herborn University Foundation, Herborn, Germany: 5-13 (2019).

SYNTHETIC SPIKES METHOD FOR PLANT MICROBIOTA ABSOLUTE QUANTITATION

ANDRZEJ TKACZ

Department of Plant Sciences, University of Oxford,
South Park Rd, OX1 3RB, Oxford, UK

THE IMPORTANCE OF PLANT SCIENCES AS A FIELD OF METHOD DEVELOPMENT

Plant science aims to improve plant yield through breeding and genetic manipulations especially focusing on plant nutrient acquisition and resistance against pathogens and abiotic factors as the weather. Plants are perfect models for studying eukaryotic genetics and hence many discoveries were first made here. To mention a few, an early genetic study on maize gave evidence that genes are physically positioned on chromosomes (*Creighton* and *McClintock*, 1931), which later allowed to find anomalies from this rule in the form of transposable elements, pieces of DNA that "jump" across genomic locations (*McClintock*, 1950). These moving DNA fragments are responsible for gene expression and phenotypic differences. It is estimated that 44% of the human genome consists of these transposable elements mostly in the form of non-coding repetitive sequences. Although approximately less than 0.05% of them are active, they cause or contribute to diseases as haemophilia and cancer (*Mills* et al., 2007). Corn, with its coloured kernels due to altered gene expression as a consequence of genetic "jumps" was a perfect study model for these discoveries. Plants allow for easy observation of their development and morphology. In the early 18th century Jean de Marian observed that *Mimosa pudica* expresses a daily leave movement even in the absence of light. This led to a further study in fruit flies that also showed a daily pattern now called circadian oscillations (*Robertson McClung*, 2006). These oscillations influence our sleep, cognitive and muscular abilities and hormone regulation. Another scientific area where plant science plays a critical role is host-microbe interactions field. In comparison to animals, plants lack defender mobile cells or adaptive immune system and hence had to develop a sophisticated innate immune and systemic signalling system to cope with the bacterial, fungal, oomycetes and insect attacks. As pathogens are responsible for a substantial crop loss (*Savary* et al., 2019), research into the plant immunity are our priority in securing food for the increasing human population. In a nutshell, plant immune system can be divided into two branches. The first branch of the immune system acts on pathogen- or microbial-associated molecular patterns (known as PAMP and MAMP) that activate immune response upon detection of well-conserved microbial proteins such as flg22, a 22 amino acid part of N-terminal part of flagellin. However, as shown by *Buscaill* et al. (2019), it is a plant role to first cleave the flagellin polymer using β-galactosidase 1 to trigger the immune response. Upon recognition, plants are not yet invaded and use salicylic acid triggered pathways to stimulate callose production and deposition to the enforcement of their cell wall.

The other immune branch acts intracellularly where pathogens release "effector" proteins to induce virulence. These effectors can be recognised by the plant host triggering a response (often cell death). At this point, the co-evolution between host and the microbe is especially pronounced as pathogens effectors are constantly evolving to escape the immune system recognition while evolving plants that can recognise these new effectors have a substantial advantage over the infected part of their population (*Jones* and *Dangl*, 2006).

Plant sciences made substantial progress with host-microbiota studies and development of methods used in this science field. In contrast to animal and especially mammalian study objects, work with plants benefit from a lack of ethical issues and the convenience and ease of propagation, crossings and seed storage as a method for preserving the host genomics population. The important species for plant scientists is *Arabidopsis thaliana*. This plant has no economic significance, however, belongs to Brassicaceae family of oilseed, cabbage and mustard. The major advantage of *A. thaliana* over other plants in genetic and host-microbiota studies is its relatively small nonrepetitive diploid genome which can be easily modified using chemical or X-ray mutagenesis. Moreover, a large pool of natural accessions (often called ecotypes) allows for studying the variations of plant response to various biotic and abiotic influence, including its interactions with environmental microbiota. A comprehensive study using two *A. thaliana* ecotypes grown in two different soils unravelled that bulk soil bacterial community is different from the rhizosphere, which in turn is different from the root compartment (*Bulgarelli* et al., 2012). Root microbiota is enriched with Proteobacteria and Actinobacteria. The study found many Streptomycetaceae (Actinobacteria) to be genuine root inhabitants, while some of the Proteobacteria being attracted purely by cellulose source as identified using wood splinters controls. This and other studies brought a new interest in plant microbiota studies. However, the methods of amplicon sequencing often employed in this science field are only able to provide a community profile snapshot without being able to even approximate their abundance. Due to the DNA isolation, PCR amplification and sequencing process any differences in the samples microbial load are completely lost. Unfortunately, standard methods as colony counting on agar media are not very useful as only a small proportion of soil and plant-associated bacteria and other microorganisms are able to grow in such conditions. There may be various reasons behind this: inadequate media nutrient status, obligate symbiosis with other organism, or simply very slow growth and the danger of being overgrown by other fast-growing species. Other methods of estimating microbial presence in environmental samples include ATP and phospholipid-derived fatty acid concentration measure, flow cytometry and qPCR (*Zhang* et al., 2017). However, they are laborious, variable and at least for now of low throughput and high cost. The importance of identifying the microbial load was clearly shown with studying gut microbial communities of Crohn's disease patients. This study, using flow cytometry unravelled that the main difference between healthy and Crohn's patient gut is the bacterial load and not the community structure (*Vandeputte* et al., 2017). This relation may be true for many other human gut diseases. The microbial load may also be a predominant factor controlling antibiotic treatment efficiency, microbial colonization and recolonization patterns and the

community structure stability. Hence a new method allowing for community profiling and load measurement is needed. In this monograph, I will present my authorship method answering these issues.

SYNTHETIC SPIKING METHOD JUSTIFICATION AND COMPARISON TO THE EXISTING METHODS

There are a few different methods to measure microbial load in environmental samples. This chapter will briefly summarize them and lists their advantages and limitations. The most common method used to measure microbial gene presence is qPCR (quantitative PCR). This method was used for example to establish microbial 16S rRNA and nifH gene (coding for nitrogenase enzyme, a key enzyme in atmospheric nitrogen conversion to ammonia) presence in the wheat rhizosphere (*Rilling* et al., 2018). However, this method can only be used on already isolated environmental DNA and hence assumes that all the DNA present in a sample will be isolated. Moreover, this method does not allow for any taxonomical identification of the microbiota. A method of flow cytometry can be used to count the microbial cells as shown in *Vandeputte* et al. (2017). Microbial cells from an environmental sample after suspension in a buffer and staining with a fluorescent dye are run through the flow cytometer machine. This method allows for very accurate measurement of the number and even the shape of the cells. The limitations are a need for laborious sample preparation including filtering samples from any debris and staining. The other problem is a need for separate flow cytometer machines to measure cells of an order of magnitude different sizes (prokaryotic vs. eukaryotic). For rich samples as soil or stool, it is not possible to taxonomically assign detected cells and hence a separated metagenomic analysis is needed. *Stämmler* et al. (2016) presented a relatively easy method to combine metagenomic with gene quantitation by adding a defined amount of exogenous bacterial cells into environmental samples. By sequencing the DNA isolated from such samples, a ratio of the number of genes detected from exogenous cells (in case of Stämmler et al., thermo- and halophilic strains) to the number of genes detected from the *in situ* microbiota (gut). The limitations are the need to *a priori* knowledge which strains are not present in the samples of interest, as the exogenous species must be different from the *in situ* microbiota, the need to culture strains of unusual growth requirements and the need to control the spiking cells number through optical density and/or colony forming units counting. Moreover, unless the cells are dead or starving they may have a variable number of 16S rRNA as cells are constantly reproducing and duplicate their DNA during mitosis before they split into two separate cells. Another problem is the lack of fungal and or other eukaryotic spiking cells. This method was shown to work well with gut samples, however, more complicated environments as soil may be challenging.

The below presented method of synthetic spiking (*Tkacz* et al., 2018) bypasses problems identified above. By adding a defined amount of synthetic taxonomical genes, the absolute amount of genes of interest as a proxy of a microbial load can be measured (16S rRNA, 18S rRNA and fungal ITS will be presented; however, the method

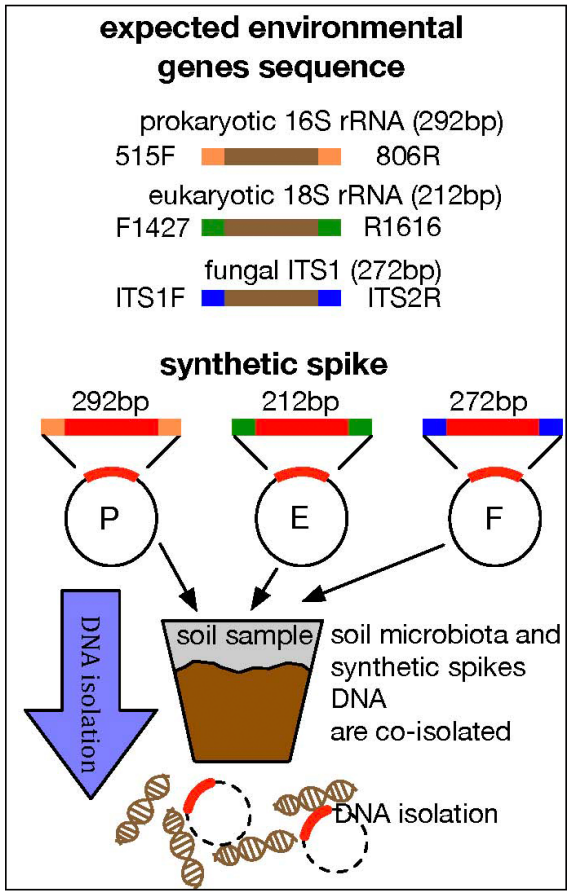


Figure 1: Synthetic spike design. Prokaryotic, eukaryotic and fungal (P, E and F) synthetic spikes in the form of plasmids (here presented as circles) were designed using primer binding site sequences, together with the length and GC content of amplicons from prokaryotic 16S rRNA (P), eukaryotic 18S rRNA (E) and fungal ITS1 (F), respectively. For P synthetic spikes the primer binding sites shown in orange, for E in green, and for F in blue (adapted from *Tkacz* et al., 2018).

can be easily expanded for the need of other genes). The spikes are added to the original sample rather than to already isolated DNA, so as the spike and the environmental DNA are co-isolated, co-amplified using a standard 16S, 18S and ITS-specific PCR (PCR where specific pairs of primers are used) and co-sequenced using high-throughput sequencing method. The method was tested using Illumina Miseq 300PE, however, any other next-generation sequencing method is suitable. The advantage of this method is that quantitation is coupled with metagenomic sequencing and hence there is no need for separate sample preparation and analysis as in the case of qPCR and flow cytometry. Moreover, the synthetic spiking method can be used to measure total prokaryotic and eukaryotic load or specific groups of microorganisms based on a selected taxonomic gene (i.e. 16S rRNA specific fragment for a given phylum) or a functional gene (i.e. nifH to measure nitrogen fixers community diversity and load). The synthetic spikes, in

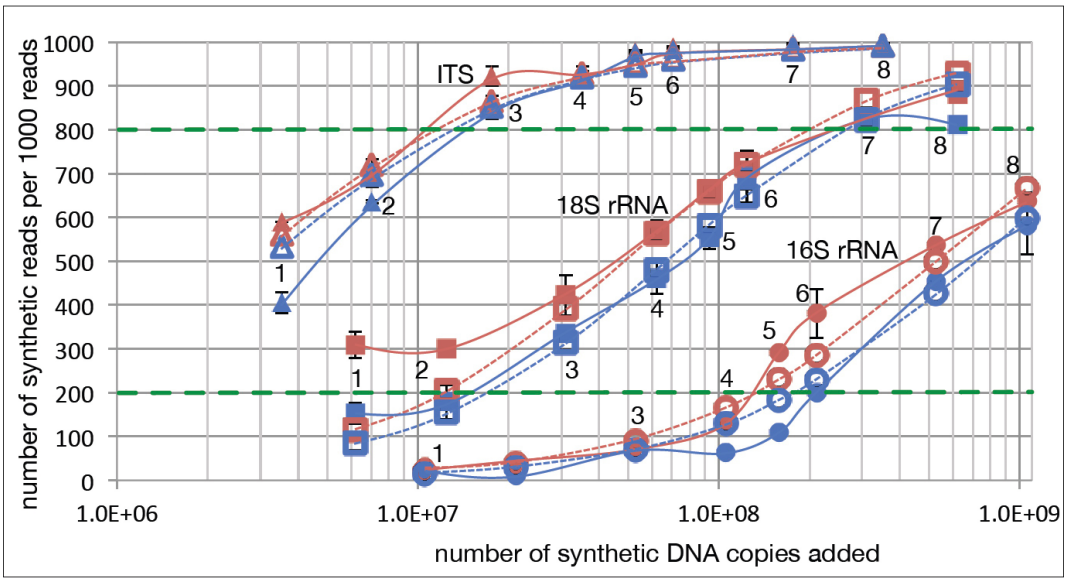


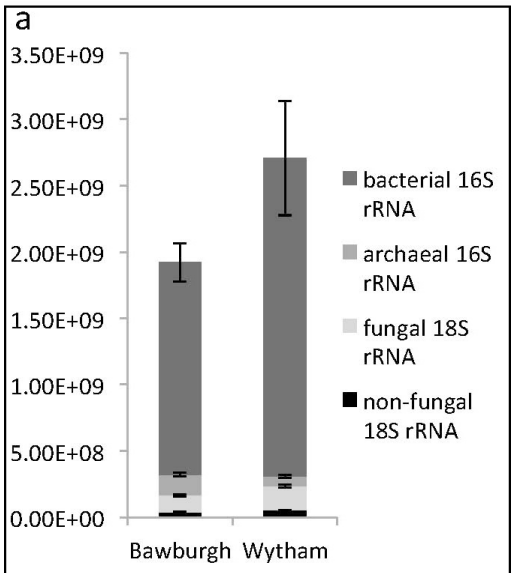
Figure 2: Model of synthetic spike addition and the number of synthetic spikes origin reads per 1000 total reads obtained in high-throughput sequencing. The number of sequencing reads of synthetic spike per 1000 total reads (Y-axis) from 16S rRNA, 18S rRNA and ITS is shown in relation to gradient levels addition to soil of prokaryotic, eukaryotic and fungal synthetic spikes (X-axis). Experimental results are shown by solid symbols and solid lines where two colours represent two soil types used as the environmental samples and model data is presented with hollow symbols and dashed lines of corresponding colours. The model shows the expected spike contribution in the sequencing output for each spike level using the averaged gene abundance for a specific soil type. Dotted green lines indicate the region with 200-800 synthetic reads per 1000 reads, where the experimental results match the model the best (adapted from *Tkacz* et al., 2018).

contrast to exogenous spiking bacterial cells, are easy to store, their amount can be easily measured using DNA quantitation methods (i.e. qubit fluorescence or nanodrop spectrophotometry) and the amount standardized between experiments (i.e. frozen synthetic spike aliquots).

SYNTHETIC SPIKES DESIGN AND APPLICATION

Synthetic spikes were designed to mimic fragments of the microbial genes of 16S rRNA, 18S rRNA and fungal ITS. The synthetic and microbial genes have the same highly conserved flanking regions to which a set of PCR primers bind, while for the synthetic spikes the region between these fragments is essentially randomly generated DNA sequence of a similar GC content as the microbial counterpart (Figure 1). Naturally, this random DNA is known, and its sequence is used to count the spikes-origin reads in the final sequencing output. The rest of the plasmid is of little importance; however, its length and sequence are used to control the accurate addition level of spikes (i.e. 1 ng of the 2666bp plasmid of a specified sequence consists of 365,572,814 copies).

Tkacz et al. (2018) have verified the synthetic spikes method accuracy by quantifying the number of bacterial



b

	Bawburgh	Wytham
bacterial 16S rRNA	83.3%	88.6%
archaeal 16S rRNA	8.1%	2.7%
fungal 18S rRNA	6.6%	6.8%
non-fungal 18S rRNA	2.0%	1.9%

Figure 3: Total soil microbial community profile obtained using synthetic spikes quantitation. a) absolute bacterial and archaeal 16S rRNA and fungal and non-fungal 18S rRNA gene abundance for each soil type (Bawburgh and Wytham) and **(b)** their relative abundance (adapted from *Tkacz* et al., 2018).

16S rRNA genes in a defined bacterial culture. A set of spikes (a 5-step gradient with 25-fold concentration levels differences) was added to *Rhizobium leguminosarum* culture, DNA isolated, PCR targeting bacterial 16S rRNA and Miseq sequencing performed. Based on the ratio of synthetic-origin to *Rhizobium*-origin sequencing reads, the number of bacterial 16S rRNA and subsequently bacterial cells was calculated. The method accuracy was calculated to be 99.3%.

Next, the spikes were added in a gradient concentration (with a difference of 100-fold between the top and the bottom gradient step) to soil samples. Tkacz and co-workers have chosen two soil types of a similar pH, but different organic carbon and nutrient content expecting differences in the microbiota community structure and the abundance. As soil is a habitat for bacteria, archaea, fungi, protists and many other eukaryotic organisms three different genes were targeted: 16S rRNA, 18S rRNA and a fungal-specific ITS region. As expected an increase in the spike addition resulted in an increase in the number of synthetic-origin sequencing reads (Figure 2). Most of the results values are close to

the expected (modelled) value further validating the method.

Absolute quantitation of soil samples allowed for an abundance comparison of prokaryotes to eukaryotes. Prokaryotic genes are about 10 times more abundant than the eukaryotes ones (Figure 3). The reason behind this is unknown. There were numerous soil microbial community studies however the real relation between these main domains of life is a yet untouched scientific field. What dictates that bacteria are more abundant than eukaryotes? Possible explanations range from their fast reproduction, small size, competitiveness to a better niche exploration. Hopefully, the future soil microbiologists will be able to shed more light on this topic.

POTENTIAL FUTURE METHOD DEVELOPMENT

Synthetic spikes as a method of a stable DNA addition to complex samples as gut or soil could be used to measure DNA isolation efficiency. There are two main DNA isolation methods used: organic extraction and solid phase extraction. Organic extraction requires lysis, phenol-chloroform separation of proteins from nucleic acids and ethanol nucleic acids precipitation. Solid phase extraction is based on silica filter DNA binding (so-called minicolumn method), its subsequent washing from any remaining contaminants (proteins, lipids) and elution using water (or buffer). Both these methods have their advantages and limitations with waste production, costs and time consumption factors. However, none of this method can isolate 100% of the DNA, especially from complex samples. Hence, ultimately it is not known what is the DNA quantity in any given sample. Synthetic spiking method allows to estimate a specific gene or genes in a sample, however, cannot establish the total DNA content. Theoretically, qPCR method could be used to measure the spike content in the isolated environmental DNA sample (synthetic spikes would need to be added to the samples prior to DNA isolation) and by comparing it to a well-defined standard of synthetic spikes DNA (qPCR amplification curve) one could establish how many copies of synthetic spikes are present in a sample after environmental DNA isolation. By a comparison of the number of synthetic spikes added to the sample to the number of them being isolated it is possible to measure the DNA isolation efficiency. The limitation is that this approach would actually be measuring the synthetic spikes isolation efficiency rather than the total DNA isolation efficiency. However, it can be assumed that samples with efficient synthetic spikes isolation rates have also their environmental DNA efficiently isolated. For the standard metagenomic PCR-based method it does not really matter what percentage of the environmental DNA has been isolated, as it can be assumed that the dominant microbial species detected in the sequencing output are actually dominant in the original environmental sample. However, if there is a focus on the so-called “rare microbiota” which involves a deep sequencing it is crucial to isolate as much of the environmental DNA as possible. A combination of synthetic spikes with qPCR would enable screening DNA samples for their usefulness in the “rare microbiota” studies.

CONCLUSIONS

Environmental ecology of microbial communities focuses on analysing prokaryotic and microbial eukaryotic profiles in complex samples as gut and soil. Due to the advance in the next generation sequencing methods, it is possible to obtain a truly deep community profile of any sample. However, a cross-samples comparison is hindered by the fact that the sequencing methods can only uncover the relative abundance of each species in comparison to the whole community. The size of the whole community is unknown. Microbial communities of a similar structure but of a different microbial load may vary in their population community stability and resistance to invasions and alterations. For example, a pathogenic communites may be more or less resistant to antibiotic treatment depending on their total abundance. Synthetic spiking method presented in this review allows measuring the microbial load using existing sequencing method without laborious and expensive additional steps. Moreover, it is theoretically possible to couple qPCR with the synthetic spiking method to measure the DNA isolation efficiency.

LITERATURE

Buscaill, P., Chandrasekar, B., Sanguankiattichai, N., Kourelis, J., Kaschani, F., Thomas, E.L., Morimoto, K., Kaiser, M., Preston, G.M., Ichinose, Y., and van der Hoorn, R.A.L.: Glycosidase and glycan polymorphism control hydrolytic release of immunogenic flagellin peptides. Science 364, 6436 (2019).

Bulgarelli, D., Rott, M., Schlaeppi, K., van Themaat, E.V.L., Ahmadinejad, N., Assenza, F., Rauf, P., Huettel, B., Reinhardt, R., Schmelzer, E., Peplies, J., Gloeckner, F.O., Amann, R., Eickhorst, T., and Schulze-Lefert, P.: Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota. Nature 488, 91-95 (2012).

Creighton, H.B. and McClintock, B.A.: Correlation of cytological and genetical crossing-over in *Zea mays*. Proc. Natl. Acad. Sci. USA. 17**,** 492-497 (1931).

Jones, J.D. and Dangl, .JL.: The plant immune system. Nature 444, 323-329. (2006).

McClintock, B.: The origin and behavior of mutable loci in maize. Proc. Natl. Acad. Sci. USA. 36, 344-355 (1950).

Mills, R.E., Bennett, E.A., Iskow, R.C., and Devine, S.E.: Which transposable elements are active in the human genome? Trends Genet. 23, 183-91 (2007).

Rilling, J.I., Acuña, J.J., Sadowsky, M.J., and Jorquera, M.A.: Putative nitrogen-fixing bacteria associated with the rhizosphere and root endosphere of wheat plants grown in an Andisol from southern Chile. Front. Microbiol. 9: 2710 (2018).

Robertson McClung, C.: Plant circadian rhythms. Plant Cell. 18, 792-803 (2006).

Savary, S., Willcocquet, L., Pethybridge, S.J., Esker, P., McRoberts, N., and Nelson, A.: The global burden of pathogens and pests on major food crops. Nat. Ecol. Evol. 3, 430-439 (2019).

Stämmler, F., Gläsner, J., Hiergeist, A., Holler, E., Weber, D., Oefner, P.J., Gessner, A., and Spang, R.: Adjusting microbiome profiles for differences in microbial load by spike-in bacteria. Microbiome 4, 28 (2016).

Tkacz, A., Hortala, M., Poole, PS.: Absolute quantitation of microbiota abundance in environmental sampels. Microbiome 6: 110 (2018).

Vandeputte, D., Kathagen, G., D'hoe, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., Wang, J., Tito, R.Y., De Commer, L., Darzi, Y., Vermeire, S., Falony, G.

and Raes, J.: Quantitative microbiome profiling links gut community variation to microbial load. Nature 551, 507-511 (2017).

Zhang, Z., Qu, Y., Li, S., Shang, K., Cai, W., Liang, Y., Li, H., Xu, M., Yin, H., and Deng, Y.: Soil bacterial quantification approaches with relative abundance reflecting the changes of taxa. Sci. Rep. 7: 4837 (2017).