Chapter 14
Strategies to Reveal Genomic Function in Natural Soil Systems

Yoshiyuki Ohtsubo, Eri Nishiyama, Yoko Ishibashi, Yuji Nagata, and Masataka Tsuda

Abstract IVET (in vivo expression technology) and STM (signature-tagged mutagenesis) are suitable methods for revealing genes that are induced in soil and are essential in soil, respectively. These methods are potentially advantageous over newer methods, such as microarray and RNA-seq, because they allow analyses of bacteria under non-sterile conditions. However, their application to non-sterile soil also presents challenges that must be overcome. In this work, we describe the difficulties we faced when using IVET to identify genes that are upregulated in *Burkholderia multivorans* ATCC 17616 in soil. We also describe the future perspectives of these and other analytical schemes, including the potential use of next-generation sequencing technologies that might allow comprehensive isolation of soil-relevant genes and also might improve the quality of research from a qualitative point of view.

Keywords Illumina • IVET • STM

14.1 Introduction

To utilize bacterial activities in natural environments, including the abilities to degrade man-made recalcitrant compounds, it is crucial to reveal the natural lives of bacteria in the environment. It is especially important to reveal their lives in the soil. It seems to be widely accepted that the transfer of pure-cultured bacteria grown in a laboratory medium into the natural ecosystem eventually results in the

Y. Ohtsubo (✉) • E. Nishiyama • Y. Ishibashi • Y. Nagata • M. Tsuda
Department of Environmental Life Sciences, Graduate School of Life Sciences,
Tohoku University, 2-1-1 Katahira, Sendai 980-8577, Japan
e-mail: yohtsubo@ige.tohoku.ac.jp

disappearance of the bacteria and in the failure to observe the expected functions of the augmented bacteria. To overcome this problem, it is important to identify genes that play pivotal roles in the survival of bacteria in the soil environments. As bacterial lives in the soil environments are not well understood, revealing such genes from the genome of each soil strain is also of particular interest from a biological point of view. In this chapter, we describe the two strategies, STM (signature-tagged mutagenesis) and IVET (in vivo expression technology), which have been used to reveal the lives of bacteria in the soil. We also describe the potential use of next-generation sequencing technologies that might aid in the studies in this field. For the general use of IVET and STM, see Rediers et al. (2005) for IVET and Mazurkiewicz et al. (2006) for STM. Also relevant is Saleh-Lakha et al. (2005) that deals with methods for microbial gene expression in soil and environmental factors affecting gene expression in soil.

14.2   Strategies to Reveal the Lives of Bacteria in a Soil Environment

One basic approach to investigate the lives of bacteria in soil environments is to inoculate the bacterial cells of interest, whose genomic sequence might be known, into soil samples. Although the augmentation of cells within a soil sample may lead to an unusual community structure in which the augmented clonal cells occupy the major part of the population, such an experimental scheme is valuable to shed light on the lives of bacteria in soil environments.

The identification of genes that play a pivotal role in natural non-sterile soil samples is compromised because of the nature of soils. First, in the soil samples, there is an immense variety of bacterial cells and eukaryotic creatures, such as fungi and possibly protozoa. This variety itself makes it impossible to isolate RNAs that originated from the augmented bacteria. Even if a sterilized soil sample is used, it is difficult to isolate RNA of sufficient quality and in sufficient amounts to be used for further analyses. This difficulty mainly arises from the presence of humic substances in the soil; these substances have physicochemical properties very similar to those of nucleic acids and are co-purified with RNA, and they interfere with the enzyme reactions and hybridization. The microarray technique was applied to analyze the expression profiles of *Pseudomonas putida* KT2440 in soils by Wang et al. who established a protocol to extract high-quality RNA from sterile soil inoculated with KT2440 (Wang et al. 2011). The other approach is to identify induced genes by adding extracts from soil to a laboratory medium and purify the RNA for further analysis. Yoder-Himes et al. used a next-generation sequencer, illumina, to conduct RNA-seq to identify genes that are expressed in response to soil extracts from in *Burkholderia cenocepacia* HI2424, which was isolated from an agricultural field (Yoder-Himes et al. 2009). Recently, genes upregulated in *Rhodococcus jostii* RHA1, a polychlorinated biphenyl degrader, during its growth in sterile soil, were reported based on microarray analysis of RNAs recovered from the sterilized soil (Iino et al. 2012).

These recent approaches have successfully identified soil-induced or soil-repressed genes. However, these studies are limited because they utilized sterile soil or soil extracts that were added to the laboratory medium. These conditions are different from the in vivo conditions, e.g., there are no competing organisms that might attack the augmented bacteria by predation or continuous production of substances with antimicrobial activities. In this regard, these new technologies allow efficient identification but fail to detect genes that might have been identified by using one of the two more traditional approaches mentioned above. These strategies, IVET and STM, enable the identification of genes that are specifically induced in soils and genes that are essential in soils, respectively. These approaches could be applied to non-sterile environments, although such applications also face difficulties and will require further improvements (see below).

IVET utilizes a positive screening scheme, in which two reporter genes play essential roles. The IVET strategy is rather complicated and there are several variations among the systems. The two major differences are the type of reporters used and the way in which the reporter genes are maintained in the bacterium, i.e., use of a plasmid vector or a specialized integration system (see Rediers et al. (2005) for the details of IVET strategies). For simplicity we here describe the IVET system that we utilized in our previous study to identify genes that are induced in the soil isolate *Burkholderia multivorans* ATCC 17616 within soil but not in a laboratory medium (Nishiyama et al. 2010). In this study, a *dapB* gene and *lacZ* gene were used as the reporter genes. *dapB* encodes an enzyme essential for the biosynthesis of lysine and diaminopimelate (DAP), and a *dapB* mutant of ATCC 17616 required the two substrates. To the *dapB* mutant, we introduced a genomic library of ATCC 17616, which was constructed in *Escherichia coli*. Each plasmid clone in the library carried a DNA region derived from ATCC 17616 fused with a tandem array of *dapB* and *lacZ*. Because of the nature of the R6K *ori*, the plasmid is not replicable in ATCC 17616, and the selection by the tetracycline-resistance gene located on the plasmid resulted in the isolation of strains, in which the plasmid was integrated in the genome by homologous recombination between the DNA region cloned in the plasmid and the corresponding DNA region in the genome (see Fig. 14.1). The IVET library was then inoculated into a soil sample, and incubated for a certain period of time for the elimination of cells not expressing the *dapB* gene (first screen). During the incubation, it was considered that the cells in which the *dapB* gene was transcribed would be able to survive or proliferate in the soil environment. After the incubation, the cell fraction was recovered and spread onto medium containing X-gal as well as DAP and lysine. We chose LacZ⁻ (i.e., white) colonies to make an output pool (second screen). Each of these clones that formed a white colony had the *dapB-lacZ* cassette in a genomic locus that is expressed in the soil environment but not in the laboratory medium. To identify the genomic regions where the plasmid was integrated, we considered two methods, one based on the retrotransfer of the integrated plasmid to *E. coli* strains, and the other based on determination of the integrated locus by sequencing using the genomic DNAs as templates, and we chose the latter (Shimoda et al. 2008). To obtain the sequencing traces we needed two unique primers that anneal to the region flanking the cloning site of the plasmid. To design such
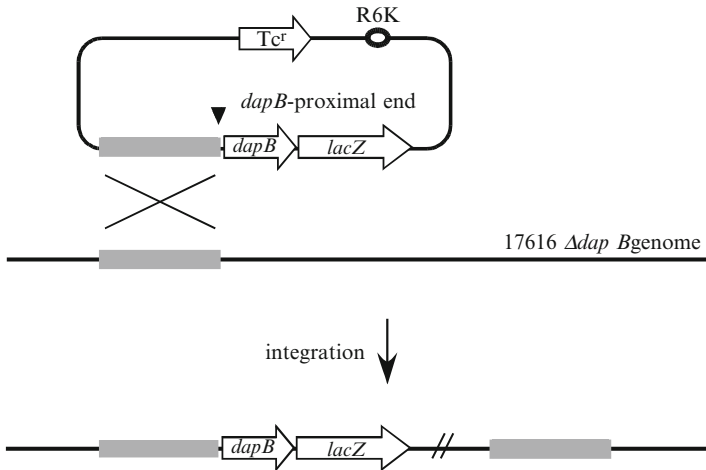
Fig. 14.1 Integration of a reporter plasmid into a genome. The *gray bar* on the plasmid represents an insert DNA derived from the strain to be analyzed. The plasmid is integrated by homologous recombination (*crossing bars*) between the insert DNA and the corresponding DNA in the target genome. In the cells, RNA polymerases are transcribing from left to right through the *dapB-proximal* end of the cloned fragment. Note that the promoter for the transcription might lie in the cloned region or might be located upstream of the cloned fragment

primers, we created a tool named "PrimerFinder," which is now available as an accessory tool of the GenomeMatcher software (Ohtsubo et al. 2008). This tool finds, from the specified region of a replicon, primers whose 3′ N-mer (i.e., 11-mer) are unique among all the DNA sequences of the specified replicons. Once we come up with two good primers to determine the two junctions, we can determine the junctions in a high-throughput manner.

One of the major problems we experienced in the application of the IVET system was the isolation of false-positive clones. The false-positive clones were clones that formed white colonies in the second screen, irrespective of the genomic location where the IVET reporters were integrated. In our preparative experiments, we constructed a negative control strain, in which the IVET plasmid was integrated to a genomic location that should not transcribe the *dapB-lacZ* cassette (the cassette was integrated in the opposite direction of a gene for cytochrome oxidase, *cox*). We also constructed a positive control strain that carries the cassette not in the opposite direction as the *cox* gene. The two strains were mixed and inoculated into the soil, and after certain time intervals the cell fraction was spread onto media containing X-gal. Ideally, the number of white colonies should reach zero after a certain period of incubation. However, white colonies were always observed at a very low, but a steady level, even after 70 days of incubation. We speculated that such white colonies appeared because a fraction of the cells had reached a dormant state after inoculation into the soil sample. There might be many places where a cell can hide itself from the other competing cells, e.g., the hollows of soil particles, leading to the survival of the cell. This speculation was supported by a finding that following the
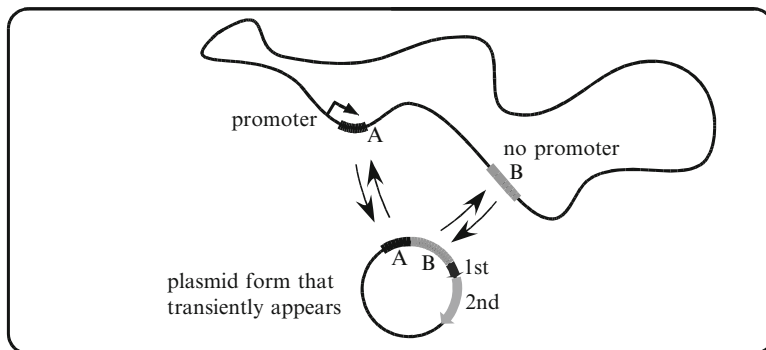
Fig. 14.2 Schematic representation of how chimeric plasmids generate false positives. If two regions (*A* and *B*) from a genome are cloned in a single plasmid, the plasmid can then be integrated into a region *A* or *B* in the genome or transiently exist in the cell in a free plasmid form. Here, it is presumed that transcription of the two reporters (designated as first and second) occurs in the cell when the plasmid is integrated into the region *A* but not into the region *B*. The integrant at region *A* can survive the soil conditions and continuously form the free plasmid. The free plasmid can integrate into region *B*, resulting in an integrant which can pass the second screen of the IVET strategy because it does not express the reporter genes. To avoid false positives, it is essential to determine both ends of the insert DNA

inoculation of the *dapB* mutant into the sterile soil, the CFU of the mutant did not decrease for a long period of time, suggesting that the *dapB* mutant did not grow but could survive in the soil.

The isolation of chimeric clones is more complicated (see Fig. 14.2). The chimeric clones are IVET clones that carry an integrated plasmid whose insert is comprised of more than one DNA region of the genome. The rate of chimeric clones of the IVET library constructed in *E. coli* is low, as exemplified by the plasmid extraction and sequencing of a fraction of clones in the library. However, the rate was significantly higher in the output pool (44 % of isolated clones). This increase in the rate of the chimeric clone was attributed to the chimeric plasmid reaching a complex equilibrium upon introduction to ATCC 17616, a condition under which it could not undergo autonomous replication. Under equilibrium, the plasmid is integrated in the genome at either of the genomic regions that are cloned or in its plasmid form, which might tentatively arise when the plasmid changes the integration location. The presence of such "free" plasmid is evident from the fact that such integrated plasmid could be retrieved to *E. coli* strains by means of conjugative cloning (Rainey et al. 1997). A chimeric IVET plasmid has more than one genomic region for integration, and if one region but not the other region transcribes the two reporters, the clone carrying such a plasmid can pass the first and second screens (see Fig. 14.2). In the soil, cells that carry the integrated plasmid in a genomic locus that transcribes *dapB-lacZ* survive and proliferate. In a very minor fraction of these cells, a free plasmid emerges and is integrated into a non-expressing locus, leading to the emergence of cells not expressing *dapB-lacZ*, which will be chosen in the second screen.

These two features, dormancy and chimeric clones, are the important features of IVET screening that should be taken into consideration to gain insights of biological importance. To overcome the difficulties, we constructed four libraries, and each library was screened in 60 independent tubes (total 240 tubes). Since the dormant cells were supposed to be recovered by chance, genomic loci that were identified from more than one tube were less likely to result from the dormant cells. We also tested whether or not the isolated clones were chimeric by obtaining sequence reads of both of the two junctions. We excluded chimeric clones and listed genomic loci that were identified from at least two independent clones, and loci that were experimentally confirmed to be induced in the soil by LacZ analysis of the cells recovered from the soil sample.

There were several difficulties in applying our IVET system to non-sterile environments. In a control experiment, the *dapB* mutant (LacZ$^-$) and a *dapB-expressing* strain (LacZ$^+$) were mixed and inoculated into the non-sterile soil sample. The colony-forming units (CFUs) of the inoculated cells decreased dramatically, and the ratio of white to blue colonies that were observed when the cell fraction recovered from the soil was plated onto an X-gal-containing agar plate, did not decrease significantly, only reaching about 1/3 in 90 days. This slower progression of competition between the *dapB* mutant and *dapB*-expressing strain indicated that a long incubation period (e.g., 5 years) is required. Although the IVET system has potential to be applied to non-sterile environments, prolonged incubation periods or advances in experimental settings are needed to stimulate the competition and to prevent the initial rapid decrease of CFU. For example, to prevent the initial rapid decrease in the CFU, cells could be inoculated into a small amount of sterile soil for a certain period of time to allow them to adapt to the soil environment before non-sterile soil is added.

It seems that the current IVET systems used thus far are limited in terms of the number of genomic loci they can identify. We spent about 2 years to determine the integration regions of 1,280 clones. Development of further analytical schemes that allow efficient identification of induced genomic loci will accelerate the IVET studies in the future (see also below).

STM is originated from a study that identified genes from *Salmonella typhimurium*, which has been implicated in the virulence to mice (Hensel et al. 1995). Basically, STM utilizes a negative screening scheme, in which transposon mutants that disappeared during the incubation or passage in a specific environment are searched, leading to identification of genes that are essential in the environment. Several modifications have been made to improve the system, resulting in several variations. The two major differences among the resulting systems are (1) the way in which each mutant is tagged with a specific DNA sequence and (2) the way to detect clones that have disappeared during the period of incubation in a given environment (Mazurkiewicz et al. 2006).

To date, except for a study of *Burkholderia vietnamiensis* G4 to identify essential genes in the rhizosphere (O'Sullivan et al. 2007), there has been no study in which STM was applied to reveal the essential genes in soil environments. In our laboratory, a *fur* gene (Yuhara et al. 2008) for the ferric uptake regulator was identified from

B. *multivorans* ATCC 17616 (our unpublished observation). However, to compile an entire list of genes essential in soil, further screening and confirmatory studies will be needed. The STM techniques will be replaced by new experimental schemes utilizing next-generation sequencers, which should identify essential genes in vivo more efficiently (see below).

14.3 Genes Found to be Induced or Essential in Soil

Table 14.1 lists studies that have identified genes that are induced or essential in soils or the rhizosphere. Although these studies identified tens or hundreds of such genes, in order to understand the function of the genes in soil, further dedicated studies will be needed. Here we describe examples of genes that were identified in the IVET screening and that were analyzed further to gain insight into their function in soil.

An IVET study of *P*. *fluorescens* Pf0-1 to identify genes important in soil identified 22 genes, including ten genes that were present in an antisense orientation relative to the overlapping protein-coding genes, and two studies followed to reveal the functional significance of the inversely oriented genes. The gene *iiv19* overlaps with the *leuA2* gene, which encodes an enzyme for leucine biosynthesis. A disruption study of *leuA2* resulted in a surprising finding, namely, that the absence of *leuA2* in soil is advantageous when leucine is added exogenously (Kim and Levy 2008).

Table 14.1 Studies pertaining to identify genes that are important in soil or rhizosphere

Strain	Strategy	Growth/incubation conditions used	Reference
Pseudomonas fluorescens SBW25	IVET	Sugar beet seeds soaked with bacteria and recovery of cells from rhizosphere	Rainey (1999)
Pseudomonas fluorescens Pf0-1	IVET	Sterilized sandy loam soil	Silby and Levy (2004)
Pseudomonas putida KT2440	IVET	Maize seeds soaked with bacteria and recovery of cells from rhizosphere	Ramos-Gonzalez et al. (2005)
Burkholderia vietnamiensis G4	STM	Pea rhizosphere	O'Sullivan et al. (2007)
Burkholderia cenocepacia HI2424/AU1054	RNA-seq	Soil medium containing soil extracts	Yoder-Himes et al. (2009)
Burkholderia multivorans ATCC 17616	IVET	Sterilized/non-sterilized brown forest field soil	Nishiyama et al. (2010)
Pseudomonas putida KT2440	Microarray	Sterilized soil in the presence or absence of 3-chlorobenzoic acid	Wang et al. (2011)
Rhodococcus jostii RHA1	Microarray	Sterilized brown forest field soil	Iino et al. (2012)

The gene *iiv8* is upregulated in soil and is antisense to the *ppk* gene that encodes polyphosphate kinase. The induction of antisense RNA encoded by *iiv8* reduced the *ppk* transcript to a level that was 1/5 of the uninduced control, suggesting a post-transcriptional mechanism. It was also suggested that precise control of polyphospate production is important for survival in the soil environment (Silby et al. 2012). As these two studies demonstrated, antisense RNAs might play significant roles in soil, suggesting that the transcriptions that generate transcripts that are complementary to protein-coding mRNA should not be ignored.

A cluster of genes (*andAcAdAbAa*) for anthranilate dioxygenase was identified in the IVET screen of *B. multivorans* ATCC 17616, and the expression level in the soil increased more than 100-fold as determined by measurement of the LacZ activities of cells recovered from the inoculated soil (Nishiyama et al. 2010). The disruption of the *andA* operon resulted in a strain that failed to proliferate in the initial period after inoculation (the initial proliferation started after one week). The expression level of the *andA* locus remained low until, after four days, it started to increase, and then it increased further after two weeks, suggesting that the *andA* expression plays a pivotal role in the initial proliferation of cells after inoculation into the soil (Nishiyama et al. 2012). In laboratory medium, the *andA* operon expression was induced by anthranilate and tryptophan. But no anthranilate or tryptophan was detected in the soil, and thus the origin of the inducer(s) and the physiological significance of the *andA* operon in soil remain a matter of speculation (Nishiyama et al. 2012).

14.4   Factors Controlling Gene Expression in Soil

It is of particular interest to identify the signals that induced the individual genes in soil (see Fig. 14.3 for potential inducers). To date, no global regulator that induces the expression of a set of genes in soil environments has been found. It seems that no soil-specific sigma factor or soil-specific transcriptional regulator exerts its effects to make a global expression profile in the soil. Rather, different independent signals present in the soil induce the respective genes.

Among the possible signal types, one of the most plausible is the low-molecular-weight signal, such as that given by organic compounds or by iron, arsenic, or other ionic forms of metals. In fact, soil extracts prepared by washing the soil with water or by organic solvents such as ethyl acetate have been shown to induce a group of genes in laboratory media (Yoder-Himes et al. 2009; Nishiyama et al. 2010). It is possible that antimicrobial agents, such as antibiotics and bacteriocins produced by co-residing bacterial or eukaryotic cells, are the inducing signals. However, it should be noted that soil is made up of soil particles that do not allow the free diffusion of chemical signals. That is, a chemical compound can be locally accumulated to reach a level sufficient to induce a gene in a bacterial cell present at that very location, but not to a level that upon extraction leads to induction in laboratory media. It is also possible that chemical compounds accumulated in the soil are the inducer.
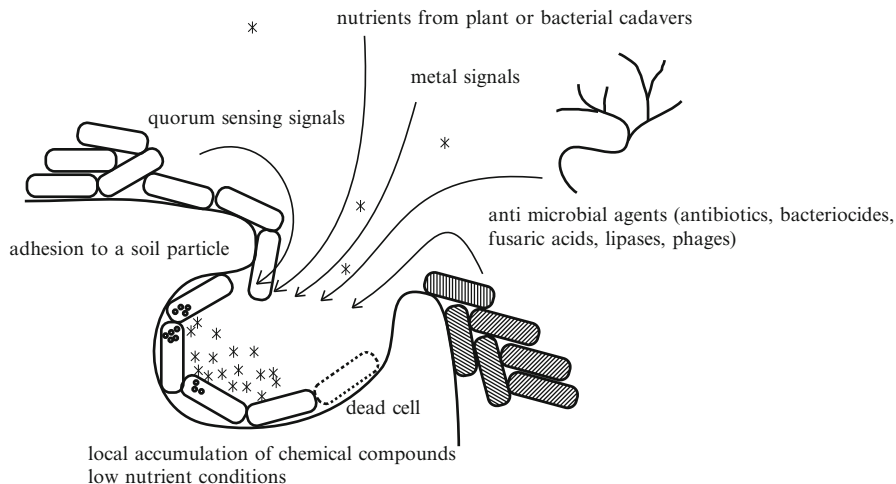
Fig. 14.3 Signals that might generate a soil-specific gene expression profile. White cells drawn in a cavity of a soil particle are the cells for analysis. These cells might be influenced by quorum-sensing signals produced by akin cells or by antimicrobial agents from fungi or other kinds of bacterial cells (hatched cells). Chemical substances could be locally accumulated (*asterisks*). Under low-nutrient conditions, particular substances such as anthranilate or tryptophan may accumulate in the cell (*dots*)

The *andA* genes in ATCC 17616, which encode anthranilate dioxygenase and are involved in anthranilate metabolism, were induced by anthranilate and tryptophan in a laboratory medium. However, neither tryptophan, which is converted to anthranilate by metabolism, nor anthranilate was detected in the soil extract. Moreover, the induction in the soil required several days of incubation, indicating that the anthranilate is not present in the soil sample but accumulated in the cell after the onset of incubation in the soil. These possibilities might account for the failure of induction by the extracts added to the laboratory medium.

Another difficulty with identifying the signal in vivo is that identification of the inducing signal in a laboratory medium might suggest but not prove that the same signal inducer induced the expression. For example, genes for fusaric acid-resistance in ATCC 17616 are upregulated in the soil, and their induction in the laboratory media requires a gene for the LysR-type transcriptional regulator, FusR, and the addition of fusaric acid (our unpublished observation). This finding in laboratory media suggested that fusaric acid is the inducer in the soil; however, the involvement of other inducers and regulatory proteins could not be excluded due to the complex nature of the soil.

Although there should be other types of signals involved in producing a soil-specific expression profile, at present only speculations can be made. It is possible that close proximity to the wall of the soil particles might be recognized by a bacterial cell to change its expression profile. It is also possible that stressful conditions that

were brought about by the poor nutrient conditions, contents of dead cells, presence of phages in the soil, and quorum-sensing molecules produced by akin cells might be responsible for generating the soil-specific expression profiles.

14.5   Future Perspectives

The TraSH method, which utilizes microarray technology, was first developed to identify essential genes in *Mycobacterium bovis* on minimal but not rich medium (Sassetti et al. 2001). Like STM, the TraSH procedure utilizes a transposon mutant library, but the identification of the transposon-inserted genes that result in the decreased fitness is done by a microarray co-hybridized with probes generated from input and output pools (Sassetti et al. 2001). In 2009, applications of three new technologies, named HITS (Gawronski et al. 2009), Tn-Seq (van Opijnen et al. 2009), and traDIS (Langridge et al. 2009), were reported. Each utilizes illumina-sequencing technology to identify the location of the transposon cassette in the massive parallel sequencing. The application of the illumina-sequencing greatly increased the efficiency of screening, making it possible to identify the insertion site and simultaneously to qualitatively assess the fitness of the mutants. Since then, several reports that utilize these technologies have been published (Smith et al. 2010; Gallagher et al. 2011; Khatiwara et al. 2012; Eckert et al. 2011), and use of the next-generation sequencing technologies will be the standards in the next decade. These technologies will be replacing the STM strategy in the future because specific tag sequences to discriminate each mutant will no longer be used.

In contrast, although it would be possible to apply the next-generation sequencing technologies to the IVET system, their use will be limited to the identification step of the insertion regions. However, to exclude chimeric clones, both junction sequences must be determined for each clone. Therefore, a new strategy must be developed to identify both junctions of each strain present in a mixture of strains.

To utilize next-generation sequencing technologies to identify soil-induced genes, the next challenge will be to augment a strain of interest into a non-sterile soil sample and prepare the RNA for illumina sequencing. As the illumina produces a huge amount of data, and the data production scale is still increasing, in the near future, it will be possible to collect a great number of sequence reads derived from the input bacteria (i.e., reads that are 100 % matched to the genomic sequence of the input bacteria) while excluding those derived from the indigenous bacteria which should dominate the whole sequence reads, so that the data can be assessed from a quantitative point of view.

The studies of microorganisms in soil are hindered by the fact that different soil samples are so different with respect to the compositions and also to the bacterial flora present in the soil. This makes it difficult to compare the results from different studies that use different soil samples. It would be possible to develop and use a formulated soil made up of known materials that are readily available, such as crushed igneous rocks and leaves of a certain kind of trees, and augment this with a

mixture of bacterial cells with known genomic sequences. The other alternative might be to decide to use a representative soil sample and share it among the laboratories. Our laboratory and those of other researchers have shared a brown forest soil sample collected from the Ehime Agricultural Experiment Station (Matsuyama, Japan). The physicochemical properties of the soil have already been reported by Wang et al. (2008).

It will also be challenging to develop an experimental scheme to recover inoculated bacterial cells from a mixture of a huge number of other kinds of cells, possibly by using the surface antigen of the input cells. The endogenous surface antigen could be used or bacteria could be engineered to express an exogenous one. The other idea is to embed the cells of the library in small permeable capsules that allow the diffusion of small molecules across the membrane, but protect the inside cells from being rapidly killed by creatures in soil. The recovery of such capsules from a soil sample would be easy, and the inside cells could be used for further analyses. For example, an RNA sample could be prepared for further analysis including RNA-seq (Mortazavi et al. 2008; Nagalakshmi et al. 2008; Yoder-Himes et al. 2009) or dRNA-seq for transcription start site analyses (Sharma et al. 2010) that utilize next-generation sequencing technologies.

References

Eckert SE, Dziva F, Chaudhuri RR, Langridge GC, Turner DJ, Pickard DJ, Maskell DJ, Thomson NR, Stevens MP (2011) Retrospective application of transposon-directed insertion site sequencing to a library of signature-tagged mini-Tn5Km2 mutants of *Escherichia coli* O157:H7 screened in cattle. J Bacteriol 193(7):1771–1776. doi:10.1128/JB.01292-10

Gallagher LA, Shendure J, Manoil C (2011) Genome-scale identification of resistance functions in *Pseudomonas aeruginosa* using Tn-seq. mBio 2(1):e00310–e00315. doi:10.1128/mBio.00315-10

Gawronski JD, Wong SM, Giannoukos G, Ward DV, Akerley BJ (2009) Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for *Haemophilus* genes required in the lung. Proc Natl Acad Sci USA 106(38):16422–16427. doi:10.1073/pnas.0906627106

Hensel M, Shea JE, Gleeson C, Jones MD, Dalton E, Holden DW (1995) Simultaneous identification of bacterial virulence genes by negative selection. Science 269(5222):400–403

Iino T, Wang Y, Miyauchi K, Kasai D, Masai E, Fujii T, Ogawa N, Fukuda M (2012) Specific gene responses of Rhodococcus jostii RHA1 during growth in soil. Appl Environ Microbiol 78(19):6954–6962. doi:10.1128/AEM.00164-12

Khatiwara A, Jiang T, Sung SS, Dawoud T, Kim JN, Bhattacharya D, Kim HB, Ricke SC, Kwon YM (2012) Genome scanning for conditionally essential genes in *Salmonella enterica* serotype typhimurium. Appl Environ Microbiol 78(9):3098–3107. doi:10.1128/AEM.06865-11

Kim W, Levy SB (2008) Increased fitness of *Pseudomonas fluorescens* Pf0-1 leucine auxotrophs in soil. Appl Environ Microbiol 74(12):3644–3651

Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G, Wain J, Parkhill J, Turner AK (2009) Simultaneous assay of every *Salmonella typhi* gene using one million transposon mutants. Genome Res 19(12):2308–2316. doi:10.1101/gr.097097.109

Mazurkiewicz P, Tang CM, Boone C, Holden DW (2006) Signature-tagged mutagenesis: barcoding mutants for genome-wide screens. Nat Rev Genet 7(12):929–939. doi:10.1038/nrg1984

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5(7):621–628. doi:10.1038/nmeth.1226

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320(5881): 1344–1349. doi:10.1126/science.1158441

Nishiyama E, Ohtsubo Y, Nagata Y, Tsuda M (2010) Identification of Burkholderia multivorans ATCC 17616 genes induced in soil environment by in vivo expression technology. Environ Microbiol 12(9):2539–2558. doi:10.1111/j.1462-2920.2010.02227.x

Nishiyama E, Ohtsubo Y, Yamamoto Y, Nagata Y, Tsuda M (2012) Pivotal role of anthranilate dioxygenase genes in the adaptation of *Burkholderia multivorans* ATCC 17616 in soil. FEMS Microbiol Lett 330(1):46–55. doi:10.1111/j.1574-6968.2012.02532.x

Ohtsubo Y, Ikeda-Ohtsubo W, Nagata Y, Tsuda M (2008) GenomeMatcher: a graphical user interface for DNA sequence comparison. BMC Bioinformatics 9:376

O'Sullivan LA, Weightman AJ, Jones TH, Marchbank AM, Tiedje JM, Mahenthiralingam E (2007) Identifying the genetic basis of ecologically and biotechnologically useful functions of the bacterium *Burkholderia vietnamiensis*. Environ Microbiol 9(4):1017–1034. doi:10.1111/j.1462-2920.2006.01228.x

Rainey PB (1999) Adaptation of *Pseudomonas fluorescens* to the plant rhizosphere. Environ Microbiol 1(3):243–257

Rainey PB, Heithoff DM, Mahan MJ (1997) Single-step conjugative cloning of bacterial gene fusions involved in microbe-host interactions. Mol Gen Genet 256(1):84–87

Ramos-Gonzalez MI, Campos MJ, Ramos JL (2005) Analysis of *Pseudomonas putida* KT2440 gene expression in the maize rhizosphere: in vivo [corrected] expression technology capture and identification of root-activated promoters. J Bacteriol 187(12):4033–4041

Rediers H, Rainey PB, Vanderleyden J, De Mot R (2005) Unraveling the secret lives of bacteria: use of in vivo expression technology and differential fluorescence induction promoter traps as tools for exploring niche-specific gene expression. Microbiol Mol Biol Rev 69(2):217–261

Saleh-Lakha S, Miller M, Campbell RG, Schneider K, Elahimanesh P, Hart MM, Trevors JT (2005) Microbial gene expression in soil: methods, applications and challenges. J Microbiol Methods 63(1):1–19. doi:10.1016/j.mimet.2005.03.007

Sassetti CM, Boyd DH, Rubin EJ (2001) Comprehensive identification of conditionally essential genes in mycobacteria. Proc Natl Acad Sci USA 98(22):12712–12717. doi:10.1073/pnas.231275498

Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermuller J, Reinhardt R, Stadler PF, Vogel J (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. Nature 464(7286):250–255. doi:10.1038/nature08756

Shimoda Y, Mitsui H, Kamimatsuse H, Minamisawa K, Nishiyama E, Ohtsubo Y, Nagata Y, Tsuda M, Shinpo S, Watanabe A, Kohara M, Yamada M, Nakamura Y, Tabata S, Sato S (2008) Construction of signature-tagged mutant library in *Mesorhizobium loti* as a powerful tool for functional genomics. DNA Res 15(5):297–308

Silby MW, Levy SB (2004) Use of in vivo expression technology to identify genes important in growth and survival of *Pseudomonas fluorescens* Pf0-1 in soil: discovery of expressed sequences with novel genetic organization. J Bacteriol 186(21):7411–7419

Silby MW, Nicoll JS, Levy SB (2012) Regulation of polyphosphate kinase production by antisense RNA in *Pseudomonas fluorescens* Pf0-1. Appl Environ Microbiol 78(12):4533–4537. doi:10.1128/AEM.07836-11

Smith AM, Heisler LE, St Onge RP, Farias-Hesson E, Wallace IM, Bodeau J, Harris AN, Perry KM, Giaever G, Pourmand N, Nislow C (2010) Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples. Nucleic Acids Res 38(13):e142. doi:10.1093/nar/gkq368

van Opijnen T, Bodi KL, Camilli A (2009) Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nat Methods 6(10):767–772. doi:10.1038/nmeth.1377

Wang Y, Shimodaira J, Miyasaka T, Morimoto S, Oomori T, Ogawa N, Fukuda M, Fujii T (2008) Detection of *bphAa* gene expression of *Rhodococcus* sp. strain RHA1 in soil using a new method of RNA preparation from soil. Biosci Biotechnol Biochem 72(3):694–701

Wang Y, Morimoto S, Ogawa N, Fujii T (2011) A survey of the cellular responses in *Pseudomonas putida* KT2440 growing in sterilized soil by microarray analysis. FEMS Microbiol Ecol 78(2):220–232. doi:10.1111/j.1574-6941.2011.01146.x

Yoder-Himes DR, Chain PS, Zhu Y, Wurtzel O, Rubin EM, Tiedje JM, Sorek R (2009) Mapping the *Burkholderia cenocepacia* niche response via high-throughput sequencing. Proc Natl Acad Sci USA 106(10):3976–3981

Yuhara S, Komatsu H, Goto H, Ohtsubo Y, Nagata Y, Tsuda M (2008) Pleiotropic roles of iron-responsive transcriptional regulator Fur in *Burkholderia multivorans*. Microbiology 154(Pt 6):1763–1774