Chapter 13

Next generation sequencing-based metagenomics for monitoring soil microbiota

Hana Yi, Department of Environmental Health, Korea University
and Jongsik Chun, School of Biological Sciences, Seoul National University

DNA sequencing is a powerful method to unravel the genetic diversity of micro-organisms in nature. In recent years, revolutionary next-generation sequencing technologies have become widely used in various microbiological disciplines, including microbial taxonomy and ecology. This chapter reviews the species concept of prokaryotes, including bacteria and Archaea, and presents the development of a comprehensive methodology for monitoring microbes in soil. Next-generation sequencing-enabled metagenomics should be useful and can be widely applied to modern microbiology and biotechnology.

Next-generation sequencing

In 1977, the chain-termination based DNA sequencing method was developed by Frederick Sanger (Sanger et al., 1977). The principle of this chain-termination method (or Sanger method) was the incorporation of dideoxynucleotide triphosphates (ddNTPs) as DNA chain terminators during the synthesis of complementary strand of template single-stranded DNA. As the ddNTPs are radioactively labelled, DNA fragments that are the result of chain termination after incorporation of ddNTPs can be detected based on one-dimensional polyacrylamide gel electrophoresis and autoradiography. The dramatic improvement of the original Sanger method was achieved by using fluorescently labelled ddNTPs and capillary electrophoresis (Smith et al., 1985; 1986). By the development of this automated Sanger sequencing method, DNA sequencing has become easier and orders of magnitude faster. The partially automated Sanger DNA sequencing method has dominated the fields of molecular biology for almost two decades and led to numerous scientific accomplishments, including the completion of the only finished-grade human genome sequence (Consortium, 2004). Despite substantial technical improvements during this period of time, the limitations of automated Sanger sequencing arose and presented a strong need for new and improved technologies for DNA sequencing with much higher throughput, such as required for sequencing large numbers of human genomes. Recent efforts have been directed towards the development of methods with a completely new basis, leaving Sanger sequencing with fewer reported incremental advances (Metzker, 2010).

Very recently, several types of high-throughput and low-cost platform for DNA sequencing methods have been developed and have made important progress in DNA sequencing (Mardis, 2008; Margulies et al., 2005; Valouev et al., 2008). The automated Sanger method is considered as a "first-generation" technology, and these newer methods are referred to as next-generation sequencing (NGS) (Pettersson et al., 2009). Currently, several NGS technologies are commercially available or about to become available, including Roche/454 (Margulies et al., 2005), Illumina/Solexa (Bentley et al., 2008), Life Technologies/APG (Valouev et al., 2008), Helicos BioSciences (Harris et al., 2008), Polonator (Shendure et al., 2005), Pacific Biosciences (Eid et al., 2009), Oxford Nanopore Technologies (Clarke et al., 2009) and Life Technologies/Ion Torrent (Rothberg et al., 2011). These new technologies employ various strategies applying multiple technological disciplines and rely on a combination of template preparation, sequencing and imaging, and genome alignment and assembly methods. One of the major advances offered by NGS is the ability to generate an enormous volume of data cheaply – in some cases in excess of 1 billion short reads per instrument run. This feature puts NGS into the new realm of experimentation such as transcriptomics, beyond just determining the order of bases (Metzker, 2010).

454 pyroseqencing

Currently, the Roche/454 pyrosequencing method dominates the NGS market together with Illumina/Solexa Genome analyzer (GA). The pyrosequencing of Roche/454 is a technology to be first introduced commercially among the next-generation sequencing methods. The pyrosequencing is a massively parallel sequencing technique based on enzymatic detection of inorganic pyrophosphate release on nucleotide incorporation (Leamon et al., 2003; Ronaghi et al., 1998). This technology employed emulsion PCR for amplification of template DNA where a single DNA template is attached to a single primer-coated bead that is then amplified to form a clonal colony

inside water droplets in an oil solution. The sequencing takes place in many picolitre-volume wells each containing a single bead and sequencing enzymes. Pyrosequencing uses luciferase to generate light for detection of the incorporation of individual nucleotides added to the nascent DNA, and the combined data are used to generate sequence read-outs (Margulies et al., 2005).

This technology provides intermediate read length and price per base compared to the conventional Sanger sequencing on one end and Illumina GA and Life Technologies SOLiD on the other (Schuster, 2008). The first version of pyrosequencing machine, called 454 Genome Sequencer (GS) 20, was released in 2004. It has been improved in the second version, 454 GS FLX, with great enhancements in terms of single-read accuracy and read length (average read length of 250 bp). The latest version of FLX series, called 454 GS FLX Titanium, generates more than 1 000 000 individual reads with improved quality of 400-500 bp in length per 10-hour instrument run (Droege and Hill, 2008; Metzker, 2010). It is currently applied to a wide variety of biological studies, such as human genetics, RNA analysis, metagenomics and ancient DNA sequencing.

Bacterial species concept and its use of genome sequence in taxonomy and metagenomics

One of the primary goals of metagenomics of the environment is to characterise the micro-organisms present in a given environmental sample as understanding the taxonomic composition of microbial communities can lead to an understanding of their ecology and function. A prokaryotic species concept is a fundamental basis of such an analysis.

A prokaryotic species is defined as a group of genetically related strains with the type strain as a centroid. A species boundary is defined by either DNA-DNA hybridisation (DDH) or 16S rRNA gene sequence similarity values. A 70% similarity level over the genome by whole genome DDH is the golden standard for species delineation (Wayne et al., 1987). The general principle of DDH requires: *i)* shearing the genomic DNA(gDNA) of the target strain and reference strains into small fragments of 1 Kb; *ii)* dissociating the double-strand gDNAs into single-strands by heating the mixture of DNA from both strains; *iii)* reannealing the fragments by subsequently decreasing the temperature. The hybrid DDH value is usually specified relative to the DDH value obtained by hybridising a reference genome with itself (Auch et al., 2010). However, the complex and time-consuming experimental procedure of this technique and the impossibility of building cumulative databases based on DDH results are the major drawbacks of this method. Thus, 16S rRNA gene has served as the primary key for phylogeny-based identification among the several thousand genes within a bacterial genome, because the amount of evolution or dissimilarity between the – highly conserved – rRNA sequences represents the variation shown by the corresponding genomes (Woese and Fox, 1977). A cutoff of 3% divergence in 16S rRNA has been used as a conservative criterion for species demarcation (Stackebrandt and Goebel, 1994; Tindall et al., 2009; Wayne et al., 1987).

In microbial molecular ecology, an operational taxonomic unit (OTU) or phylotype often corresponds to a prokaryotic species, which is defined as a group of organisms with high (≥97%) 16S rRNA gene sequence homology. The identification of new bacterial isolates also widely relies on the 16S rRNA gene sequence homology analysis by comparison with existing sequences in the reference databases. Because of the experimental simplicity and the availability of public databases of 16S rRNA gene

sequences, the use of this gene as a single marker for species circumscription has been well received, and it will be argued below that useful metragenomics data can be based on the study of 16S RNA. However, being a highly conserved molecule, the 16S rRNA gene does not always provide sufficient resolution at species and strain level (Konstantinidis et al., 2006). Moreover, single gene-based phylogeny may cause problems because of the possibility of horizontal gene transfer and intra-genomic heterogeneity of multiple copies of the genes (Rajendhran and Gunasekaran, 2011). The experimental difficulty of DDH and the lack of resolution of 16S rRNA gene sequence within species have raised the demand for a better method for species delineation (Stackebrandt et al., 2002).

Now, in the NGS era, in which high-quality genome sequence can be analysed easily and can be compared with other genomes in the public databases, average nucleotide identity (ANI) value between a given pair of genomes has been recognised as a simple and effective way to reconcile the genomic information with the current prokaryotic species concept (Goris et al., 2007; Konstantinidis and Tiedje, 2005). The inter-genomic distances are calculated from fully or partially sequenced genomes after cutting them into small pieces *in silico* (e.g. 1020 bp-long). Then, high-scoring segment pairs (HSPs) between two genome sequences are determined using BLAST algorithm (Altschul et al., 1997; Goris et al., 2007), or maximally unique matches (MUMs) between genome sequences are determined using MUMmer, an ultra-rapid aligning tool (Kurtz et al., 2004; Richter and Rossello-Mora, 2009). The ANI is then calculated from the sets of HSPs or MUMs. The comparative efforts undertaken to evaluate the ANI led to ascertain that the ANI reflects the degree of evolutionary distance between the compared genomes, and a value of 94-96% identity represents the DDH boundary of 70% (Auch et al., 2010; Goris et al., 2007; Konstantinidis and Tiedje, 2005; Richter and Rossello-Mora, 2009). The cases of using the ANI as a substitution of the DDH are beginning to increase in taxonomic studies (Vanlaere et al., 2009; Yi et al., 2012).

Microbial community analysis: Conventional methods

It is generally known among microbiologists that there is a huge potential of prokaryotic diversity made up of hitherto uncultured micro-organisms (Pace, 1997; Ward et al., 1990). Molecular techniques directed toward analysing the community composition of environmental samples indicate that hitherto classified prokaryotic species account for only the tip of the iceberg, considering the huge number (estimated as 4-6 x 10^{30}) of undiscovered prokaryotes present on Earth (Whitman et al., 1998). Usually, profiles of microbial communities in environments have been surveyed using genetic fingerprinting methods. Genetic fingerprinting is a DNA-based technique which generates a fingerprint, the barcode-like DNA fragment pattern. This is a direct analysis of whole genomes extracted from environments or PCR products of selected genes amplified from environmental DNA, based on either sequence polymorphism or length polymorphism. These techniques include denaturing gradient gel electrophoresis (DGGE), temperature gradient gel electrophoresis (TGGE), terminal restriction fragment length polymorphism analysis (T-RFLP), single-strand conformation polymorphism (SSCP), random amplified polymorphic DNA (RAPD), ribosomal intergenic spacer analysis (RISA), length heterogeneity PCR (LH-PCR), amplified ribosomal DNA restriction analysis (ARDRA) and DNA microarrays. In general, genetic fingerprinting techniques are simple and rapid, and allow simultaneous analyses of a large number of multiple samples. The "fingerprints" from different samples are then compared using computer-assisted cluster analysis and community relationships or differences between

microbial communities are inferred (Rastogi and Sani, 2011). However, fingerprinting approaches do not provide direct taxonomic identities of the members comprising the microbial community. Building up a comparable database is also impossible for fingerprinting-based methodology due to the variability of fingerprinting patterns depending on the gel-electrophoresis conditions.

Microbial community analysis: Metagenomics

Thanks to recent technological advancements, methods for the elucidation of microbial community structures have shifted from indirect methods, such as DGGE, T-RFLP and DNA microarrays, to direct methods called metagenomics (Rondon et al., 2000; Schmidt et al., 1991). Metagenomics is a study of collective set of genetic materials extracted directly from environmental samples, and does not rely on cultivation or prior knowledge of the microbial communities (Riesenfeld et al., 2004). Thus, it is a powerful tool to unravel environmental genetic diversity without potential biases resulting from culturing or isolation. Metagenomics is also known by other names, such as environmental genomics or community genomics, or microbial ecogenomics (Rastogi and Sani, 2011). The two major interests of metagenomics are which organisms are present and what metabolic processes are possible in the community (Allen and Banfield, 2005). The former is surveyed mainly based on 16S rRNA gene profiling, the prevalent marker gene for identification of prokaryotic species (Weisburg et al., 1991). Metagenomic investigations have been conducted in several environments, ranging from the oceans to soil, the phyllosphere and acid mine drainage, and have provided access to phylogenetic and functional diversity of uncultured micro-organisms (Handelsman, 2004).

Several major technical limitations have long been in existence with respect to metagenomics. PCR was usually used in metagenomics to selectively amplify target genes and then cloned into vectors for sequencing (Lane et al., 1985). This approach could amplify a minute amount of target genes from the bulk DNA to a reasonable quantity for analysis, but this analysis is subject to PCR-inherent bias (Polz and Cavanaugh, 1998) and thus may not reflect actual microbial community structure. By the advances of meta-strategies in biotechnology and bioinformatics, the need for PCR can be avoided by adopting shotgun sequencing into metagenomics (Breitbart et al., 2002; Tyson et al., 2004). This was feasible by using randomly sheared environmental DNA as it is for insert to be sequences, but still the potential bias imposed by cloning remained as a significant concern in shotgun metagenomics (Handelsman, 2004).

As described above, NGS methods such as Roche 454 pyrosequencing have brought a revolution in metagenomics not only by producing a large amount of data at a low cost, but also by excluding time-consuming and bias-imposing step such as clone library construction.

For the purpose of collecting metagenomics data, DNA is extracted from an entire microbial community, and a target region flanked by highly conserved primers is amplified by PCR before sequencing. This generates a mixture of amplicons, in which every read stems from a homologous region, and the sequence variation between the reads reflects the phylogenetic diversity in the community (Quince et al., 2009). Usually, the hypervarialble regions of 16S rRNA gene sequences are used for the target of pyrosequencing. The produced sequences are short (400~500 bp), but provide useful phylogenetic information. For example, investigation on the spatial changes in soil bacterial communities was explored using 88 soil samples and a massive bar-coded pyrosequencing technique (Lauber et al., 2009). The V1 and V2 hypervariable region of

16S rRNA genes was the target of sequencing. The results demonstrated that soil bacterial communities contain a large number of microbial species, implying extreme diversity; at least 1 000 species per soil sample. A large "rare biosphere" represented by an enormous number of low-abundance unique taxa also supports this finding. Such studies highlight the importance of large-scale sequencing techniques in investigating the highly diverse soil microbial communities (Rastogi and Sani, 2011). Now, this kind of microbial metagenomic sequencing data itself have become generally affordable and researchers are flooded by an unprecedented amount of DNA sequence data from various environments (Huber et al., 2007; Jones et al., 2009; Warnecke et al., 2007; Wegley et al., 2007).

Soil metagenomics: Practical applications

Phytoremediation, which is the use of plants to clean up environmental pollution, has received much attention as a promising method for the removal of metal pollutants in soils (Cherian and Oliveira, 2005; Van Aken, 2008). Phytoremediation is a cost-effective and environmentally friendly approach compared to other environmentally invasive, expensive and inefficient clean-up technologies (Van Aken, 2008). A number of plant species are capable of high-level organic compound degradation or heavy metal hyperaccumulation. However, slow rates of removal and incomplete metabolism have restricted the application of phytoremediation in the field (Van Aken, 2008). Thus, genetically engineered plants that exhibit enhanced performance with respect to the metabolism of toxic compounds have been developed by the over-expression and/or introduction of genes from other organisms (Doty et al., 2007; French et al., 1999). Engineered poplars have greatly increased the possibility of the practical application of phytoremediation. However, this technology is still in the developmental stage, with the field testing of transgenic plants for phytoremediation being very limited. The major obstacle is biosafety concerns, because the potential unwanted effects of genetically modified organisms are not fully understood.

One of the most postulated potential unwanted effects of genetically modified (GM) plants is alteration to the structure of indigenous microbial communities. Micro-organisms have an important role in regulating soil conditions (Wolfenbarger and Phifer, 2000). Soil micro-organisms are in charge of the global cycling of organic and inorganic matter. A number of microbes decompose organic matter into forms useful to the rest of the organisms in the soil food web, and can break down pesticides and pollutants in soil. Soil microbes perform important services related to water dynamics, nutrient cycling and disease suppression. They also produce substances that constitute the soil structure (Conrad, 1996). Thus, alteration in the diversity or activity of microbial communities may have adverse effects on soil ecology (Kennedy and Smith, 1995), and understanding how GM plants, and plants in general, might alter the soil microbial community is of great interest.

The effect of GM plants on soil microbial communities remains highly controversial. Several studies have reported that microbial communities are clearly altered by engineered plants (Bruce et al., 2007; Donegan et al., 1999; Gyamfi et al., 2002; LeBlanc et al., 2007; Lee et al., 2011; Siciliano and Germida, 1999; Smalla et al., 2001). In contrast, other studies have shown that the associated changes in microbial communities with engineered plants are statistically insignificant (Dunfield and Germida, 2004; Heuer et al., 2002; Kim et al., 2008; Lottmann et al., 2000) or very minor (Di Giovanni et al., 1999; Donegan et al., 1995, 1999; Dunfield and Germida, 2003;

Griffiths et al., 2000; Gyamfi et al., 2002; Jain et al., 2010; Lukow et al., 2000; Schmalenberger and Tebbe, 2002). Most of these studies have used non-sequencing based methods, such as community-level physiological profiles (CLPPs), fatty acid methyl ester (FAME), DGGE and T-RFLP. These techniques are useful for evaluating differences in overall community structure, but these fingerprinting methods are limited in their capacity to detect minor changes and the components of these changes. In addition, the number of clone sequences (≤100 sequences per sample) surveyed in a few studies (Kim et al., 2008; LeBlanc et al., 2007; Lee et al., 2011) is insufficient to determine overall community profiles.

Thus, to evaluate the effect of GM plant use on soil microbial communities, extensive sequencing-based community analysis was conducted, while controlling the influence of plant clonality, plant age, soil condition and harvesting season (Hur et al., 2011). The rhizosphere soils of GM and wild type (WT) poplars at a range of growth stages (i.e. rhizosphere of 1.5-, 2.5- and 3-year-old poplars) were sampled together with non-planted contaminated soil, and the microbial community structure was investigated by pyrosequencing the V3 region of prokaryotic 16S rRNA gene. Based on the results of DNA pyrosequencing, poplar type and growth stages were associated with directional changes in the structure of the microbial community. In detail, for both GM and WT poplars, the microbial community of poplars started separating from that of the control soil in the early stage of poplar cultivation (1.5 years), advanced to the middle-stage group (2.5 years), and finally reached the late-stage group (3 years), the composition of which was very different from that of the contaminated soil community. However, the rate of microbial community change was slower in WT poplars than in GM poplars. This phenomenon possibly occurs because of the more active metal uptake ability of GM poplars compared to WT poplars, which resulted in faster changes in the soil environment, and hence the microbial habitat. In conclusion, the shift in the microbial community structure to the late stage was driven faster by the effect of GM phytoremediation than WT phytoremediation. The results of the study demonstrated the superiority of NGS-based technique over traditional risk assessment approaches in the aspect of capacity to detect minor changes and the components of these changes. The next-generation sequencing-enabled metagenomics should be useful and can be widely applied to modern microbiology and bio-technology.

Conclusion

The NGS techniques, coupled with metagenomic analysis, has opened up a new era in the study of microbial diversity with direct access to the indigenous microbial communities in the environments. The superiority of NGS-metagenomics over conventional DNA fingerprinting or Sanger-metagenomics is evident from numerous microbial diversity studies. This NGS-metagenomics also provides further research strategies at the molecular level, such as gene-level functional analysis and gene expression analysis. In a near feature, this NGS-metagenomics will be able to be used as a universal diagnostic tool also in clinical bacterial or viral samples. The new NGS-enabled diagnosis requires no prior knowledge of the host or pathogen, and thus will expedite the entire process of novel pathogen discovery, identification, pathogen genome sequencing and the development of more routine assays.

Because the NGS techniques are still rapidly evolving, researchers continue to meet challenges in fully optimising NGS platforms as well as in analysing and managing data. Many technological developments are focusing on the sample-preparation protocols,

sequencing-library construction protocol, the quality and quantity of sequencing reads, and the analysis of massive data. One of the most challenging parts of those is developing novel algorithms and bioinformatic tools that scale with the tremendous amount of short reads generated through NGS-metagenomics. As the NGS technologies are producing a tsunami of data, the bioinformatics community needs to act quickly to keep up to pace with it. Particularly for NGS-metagenomics, efforts should be made to prepare tools for error-free estimation of species diversity and gene family frequency, tools for comparative metagenomics and tools for removing 16S rRNA chimeras.

NGS-metagenomics is useful and can be widely applied to modern microbiology and biotechnology. It has the potential to answer fundamental biological questions. The current progress toward understanding the uncultured bacteria, archaea and viruses through NGS-metagenomic analyses will lead to the comprehension of the genetic diversity, population structure and ecological function of complex microbial assemblages in the environments.

References

Allen, E.E. and J.F. Banfield (2005), "Community genomics in microbial ecology and evolution", *Nature Reviews of Microbiology*, No. 3, pp. 489-498.

Altschul, S.F., et al. (1997), "Gapped BLAST and PSI-BLAST: A new generation of protein database search programs", *Nucleic Acids Research*, No. 25, pp. 3 389-3 402.

Auch, A.F., et al. (2010), "Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison", *Standards in Genomic Science*, No. 2, pp. 117-134.

Bentley, D.R., et al. (2008), "Accurate whole human genome sequencing using reversible terminator chemistry", *Nature*, No. 456, pp. 53-59.

Breitbart, M., et al. (2002), "Genomic analysis of uncultured marine viral communities", *Proceedings of the National Academy of Sciences of the United States of America*, No. 99, pp. 14 250-14 255.

Bruce, N.C., et al. (2007), "Impact of transgenic tobacco on trinitrotoluene (TNT) contaminated soil community", *Environmental Science & Technology*, No. 41, pp. 5 854-5 861.

Cherian, S. and M.M. Oliveira (2005), "Transgenic plants in phytoremediation: Recent advances and new possibilities", *Environmental Science & Technology*, No. 39, pp. 9 377-9 390.

Clarke, J., et al. (2009), "Continuous base identification for single-molecule nanopore DNA sequencing", *Nature Nanotechnology*, No. 4, pp. 265-270.

Conrad, R. (1996), "Soil micro-organisms as controllers of atmospheric trace gases (H2, CO, CH4, OCS, N2O, and NO)", *Microbiological Reviews*, No. 60, pp. 609-640.

Consortium, I.H.G. (2004), "Finishing the euchromatic sequence of the human genome", *Nature*, No. 431, pp. 931-945.

Di Giovanni, G.D., et al. (1999), "Comparison of parental and transgenic alfalfa rhizosphere bacterial communities using biolog GN metabolic fingerprinting and enterobacterial repetitive intergenic consensus sequence-PCR (ERIC-PCR)", *Microbial Ecology*, No. 37, pp. 129-139.

Donegan, K.K., et al. (1999), "A field study with genetically engineered alfalfa inoculated with recombinant *Sinorhizobium meliloti*: Effects on the soil ecosystem", *Journal of Applied Ecology*, No. 36, pp. 920-936.

Donegan, K.K., et al. (1995), "Changes in levels, species and DNA fingerprints of soil-micro-organisms associated with cotton expressing the *Bacillus thuringiensis* var. Kurstaki endotoxin", *Applied Soil Ecology*, No. 2, pp. 111-124.

Doty, S.L., et al. (2007), "Enhanced phytoremediation of volatile environmental pollutants with transgenic trees", *Proceedings of the National Academy of Sciences of the United States of America*, No. 104, pp. 16 816-16 821.

Droege, M. and B. Hill (2008), "The genome sequencer FLX system – longer reads, more applications, straight forward bioinformatics and more complete data sets", *Journal of Biotechnology*, No. 136, pp. 3-10.

Dunfield, K.E. and J.J. Germida (2004), "Impact of genetically modified crops on soil- and plant-associated microbial communities", *Journal of Environmental Quality*, No. 33, pp. 806-815.

Dunfield, K.E. and J.J. Germida (2003), "Seasonal changes in the rhizosphere microbial communities associated with field-grown genetically modified canola (*Brassica napus*)", *Applied and Environmental Microbiology*, No. 69, pp. 7 310-7 318.

Eid, J., et al. (2009), "Real-time DNA sequencing from single polymerase molecules", *Science*, No. 323, pp. 133-138.

French, C.E., et al. (1999), "Biodegradation of explosives by transgenic plants expressing pentaerythritol tetranitrate reductase", *Nature Biotechnology*, No. 17, pp. 491-494.

Goris, J., et al. (2007), "DNA-DNA hybridization values and their relationship to whole-genome sequence similarities", *International Journal of Systematic and Evolutionary Microbiology*, No. 57, pp. 81-91.

Griffiths, B.S., I.E. Geoghegan and W.M. Robertson (2000), "Testing genetically engineered potato, producing the lectins GNA and Con A, on non-target soil organisms and processes", *Journal of Applied Ecology*, No. 37, pp. 159-170.

Gyamfi, S., et al. (2002), "Effects of transgenic glufosinate-tolerant oilseed rape (*Brassica napus*) and the associated herbicide application on eubacterial and pseudomonas communities in the rhizosphere", *FEMS Microbiology Ecology*, No. 1, pp. 181-190.

Handelsman, J. (2004), "Metagenomics: Application of genomics to uncultured micro-organisms", *Microbiology and Molecular Biology Reviews*, No. 68, pp. 669-685.

Harris, T.D., et al. (2008), "Single-molecule DNA sequencing of a viral genome", *Science*, No. 320, pp. 106-109.

Heuer, H., et al. (2002), "Effects of T4 lysozyme release from transgenic potato roots on bacterial rhizosphere communities are negligible relative to natural factors", *Applied and Environmental Microbiology*, No. 68, pp. 1 325-1 335.

Huber, J.A., et al. (2007), "Microbial population structures in the deep marine biosphere", *Science*, No. 318, pp. 97-100.

Hur, M., et al. (2011), "Effect of genetically modified poplars on soil microbial communities during the phytoremediation of waste mine tailings", *Applied and Environmental Microbiology*, No. 77, pp. 7 611-7 619.

Jain, R.K., et al. (2010), "A case study for assessment of microbial community dynamics in genetically modified Bt cotton crop fields", *Current Microbiology*, No. 61, pp. 118-124.

Jones, R.T., et al. (2009), "A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses", *ISME Journal*, No. 3, pp. 442-453.

Kennedy, A.C. and K.L. Smith (1995), "Soil microbial diversity and the sustainability of agricultural soils", *Plant Soil*, No. 170, pp. 75-86.

Kim, M.C., et al. (2008), "Molecular analysis of bacterial community structures in paddy soils for environmental risk assessment with two varieties of genetically modified rice, Iksan 483 and Milyang 204", *Journal of Microbiology and Biotechnology*, No. 18, pp. 207-218.

Konstantinidis, K.T. and J.M. Tiedje (2005), "Genomic insights that advance the species definition for prokaryotes", *Proceedings of the National Academy of Sciences of the United States of America*, No. 102, pp. 2 567-2 572.

Konstantinidis, K.T., A. Ramette and J.M. Tiedje (2006), "The bacterial species definition in the genomic era", *Philosophical Transactions of the Royal Society of London B Biological Sciences*, No. 361, pp. 1 929-1 940.

Kurtz, S., et al. (2004), "Versatile and open software for comparing large genomes", *Genome Biology*, No. 5.

Lane, D.J., et al. (1985), "Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses", *Proceedings of the National Academy of Sciences of the United States of America*, No. 82, pp. 6 955-6 959.

Lauber, C.L., et al. (2009), "Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale", *Applied and Environmental Microbiology*, No. 75, pp. 5 111-5 120.

Leamon, J.H., et al. (2003), "A massively parallel PicoTiterPlate based platform for discrete picoliter-scale polymerase chain reactions", *Electrophoresis*, No. 24, pp. 3769-3777.

LeBlanc, P.M., R.C. Hamelin and M. Filion (2007), "Alteration of soil rhizosphere communities following genetic transformation of white spruce", *Applied and Environmental Microbiology*, No. 73, pp. 4 128-4 134.

Lee, Y.E., et al. (2011), "Effects of field-grown genetically modified zoysia grass on bacterial community structure", *Journal of Microbiolgy and Biotechnology*, No. 21, pp. 333-340.

Lottmann, J., et al. (2000), "Establishment of introduced antagonistic bacteria in the rhizosphere of transgenic potatoes and their effect on the bacterial community", *FEMS Microbiology Ecology*, No. 33, pp. 41-49.

Lukow, T., P.F. Dunfield and W. Liesack (2000), "Use of the T-RFLP technique to assess spatial and temporal changes in the bacterial community structure within an agricultural soil planted with transgenic and non-transgenic potato plants", *FEMS Microbiology Ecology*, No. 32, pp. 241-247.

Mardis, E.R. (2008), "Next-generation DNA sequencing methods", *Annual Reviews of Genomics and Human Genetics*, No. 9, pp. 387-402.

Margulies, M., et al. (2005), "Genome sequencing in microfabricated high-density picolitre reactors", *Nature*, No. 437, pp. 376-380.

Metzker, M.L. (2010), "Sequencing technologies – the next generation", *Nature Reviews Genetics*, No. 11, pp. 31-46.

Pace, N.R. (1997), "A molecular view of microbial diversity and the biosphere", *Science*, No. 276, pp. 734-740.

Pettersson, E., J. Lundeberg and A. Ahmadian (2009), "Generations of sequencing technologies", *Genomics*, No. 93, pp. 105-111.

Polz, M.F. and C.M. Cavanaugh (1998), "Bias in template-to-product ratios in multitemplate PCR", *Applied and Environmental Microbiology*, No. 64, pp. 3 724-3 730.

Quince, C., et al. (2009), "Accurate determination of microbial diversity from 454 pyrosequencing data", *Nature Methods*, No. 6, pp. 639-641.

Rajendhran, J. and P. Gunasekaran (2011), "Microbial phylogeny and diversity: Small subunit ribosomal RNA sequence analysis and beyond", *Microbiological Research*, No. 166, pp. 99-110.

Rastogi, G. and R.K. Sani (2011), "Molecular techniques to assess microbial community structure, function, and dynamics in the environment", in: I. Ahmad, F. Ahmad and J. Pichtel (eds.), *Microbes and Microbial Technology: Agricultural and Environmental Applicaions*, Springer.

Richter, M. and R. Rossello-Mora (2009), "Shifting the genomic gold standard for the prokaryotic species definition", *Proceedings of the National Academy of Sciences of the United States of America*, No. 106, pp. 19 126-19 131.

Riesenfeld, C.S., P.D. Schloss and J. Handelsman (2004), "Metagenomics: Genomic analysis of microbial communities", *Annual Reviews of Genetics*, No. 38, pp. 525-552.

Ronaghi, M., M. Uhlen and P. Nyren (1998), "A sequencing method based on real-time pyrophosphate", *Science*, No. 281, pp. 363-365.

Rondon, M.R., et al. (2000), "Cloning the soil metagenome: A strategy for accessing the genetic and functional diversity of uncultured micro-organisms", *Applied and Environmental Microbiology*, No. 66, pp. 2 541-2 547.

Rothberg, J.M., et al. (2011), "An integrated semiconductor device enabling non-optical genome sequencing", *Nature*, No. 475, pp. 348-352.

Sanger, F., S. Nicklen and A.R. Coulson (1977), "DNA sequencing with chain-terminating inhibitors", *Proceedings of the National Academy of Sciences of the United States of America*, No. 74, pp. 5 463-5 467.

Schmalenberger, A. and C.C. Tebbe (2002), "Bacterial community composition in the rhizosphere of a transgenic, herbicide-resistant maize (Zea mays) and comparison to its non-transgenic cultivar bosphore", *FEMS Microbiology Ecology*, No. 40, pp. 29-37.

Schmidt, T.M., E.F. DeLong and N.R. Pace (1991), "Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing", *Journal of Bacteriology*, No. 173, pp. 4 371-4 378.

Schuster, S.C. (2008), "Next-generation sequencing transforms today's biology", *Nature Methods*, No. 5, pp. 16-18.

Shendure, J., et al. (2005), "Accurate multiplex polony sequencing of an evolved bacterial genome", *Science*, No. 309, pp. 1 728-1 732.

Siciliano, S.D. and J.J. Germida (1999), "Enhanced phytoremediation of chlorobenzoates in rhizosphere soil", *Soil Biology & Biochemistry*, No. 31, pp. 299-305.

Smalla, K., et al. (2001), "Bulk and rhizosphere soil bacterial communities studied by denaturing gradient gel electrophoresis: Plant-dependent enrichment and seasonal shifts revealed", *Applied and Environmental Microbiology*, No. 67, pp. 4 742-4 751.

Smith, L.M., et al. (1986), "Fluorescence detection in automated DNA sequence analysis", *Nature*, No. 321, pp. 674-679.

Smith, L.M., et al. (1985), "The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: Synthesis of fluorescent DNA primers for use in DNA sequence analysis", *Nucleic Acids Research*, No. 13, pp. 2 399-2 412.

Stackebrandt, E. and B.M. Goebel (1994), "A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology", *International Journal of Systematic Bacteriology*, No. 44, pp. 846-849.

Stackebrandt, E., et al. (2002), "Report of the *Ad Hoc* Committee for the Re-evaluation of the Species Definition in Bacteriology", *International Journal of Systematic and Evolutionary Microbiology*, No. 52, pp. 1 043-1 047.

Tindall, B.J., et al. (2009), "Notes on the characterization of prokaryote strains for taxonomic purposes", *International Journal of Systematic and Evolutionary Microbiology*, published online ahead of print, http://dx.doi.org/10.1099/ijs.0.016949-0.

Tyson, G.W., et al. (2004), "Community structure and metabolism through reconstruction of microbial genomes from the environment", *Nature*, No. 428, pp. 37-43.

Valouev, A., et al. (2008), "A high-resolution, nucleosome position map of *C. Elegans* reveals a lack of universal sequence-dictated positioning", *Genome Research*, No. 18, pp. 1 051-1 063.

Van Aken, B. (2008), "Transgenic plants for phytoremediation: Helping nature to clean up environmental pollution", *Trends in Biotechnology*, No. 26, pp. 225-227.

Vanlaere, E., et al. (2009), "Taxon K, a complex within the *Burkholderia Cepacia* complex, comprises at least two novel species, *Burkholderia Contaminans* sp. nov. and *Burkholderia lata* sp. Nov", *International Journal of Systematic and Evolutionary Microbiology*, No. 59, pp. 102-111.

Ward, D.M., R. Weller and M.M. Bateson (1990), "16S rRNA sequences reveal numerous uncultured micro-organisms in a natural community", *Nature*, No. 345, pp. 63-65.

Warnecke, F., et al. (2007), "Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite", *Nature*, No. 450, pp. 560-565.

Wayne, L.G., et al. (1987), "Report of the *Ad Hoc* Committee on Reconciliation of Approaches to Bacterial Systematics", *International Journal of Systematic Bacteriology*, No. 37, pp. 463-464.

Wegley, L., et al. (2007), "Metagenomic analysis of the microbial community associated with the coral porites astreoides", *Environmental Microbiology*, No. 9, pp. 2 707-2 719.

Weisburg, W.G., et al. (1991), "16S ribosomal DNA amplification for phylogenetic study", *Journal of Bacteriology*, No. 173, pp. 697-703.

Whitman, W.B., D.C. Coleman and W.J. Wiebe (1998), "Prokaryotes: The unseen majority", *Proceedings of the National Academy of Sciences of the United States of America*, No. 95, pp. 6 578-6 583.

Woese, C.R. and G.E. Fox (1977), "Phylogenetic structure of the prokaryotic domain: The primary kingdoms", *Proceedings of the National Academy of Sciences of the United States of America*, No. 74, pp. 5 088-5 090.

Wolfenbarger, L.L. and P.R. Phifer (2000), "The ecological risks and benefits of genetically engineered plants", *Science*, No. 290, pp. 2 088-2 093.

Yi, H., et al. (2012), "Comparative genomics of *Neisseria Weaveri* clarifies the taxonomy of this species and identifies genetic determinants that may be associated with virulence", *FEMS Microbiology Letters*, No. 328, pp. 100-105.