ISSN 0367-6315
Korean J. Soil Sci. Fert. 45(6), 1073-1085 (2012)
http://dx.doi.org/10.7745/KJSSF.2012.45.6.1073

총설

토양미생물 생태 연구를 위한 증폭 파이로시퀀싱 기법의 응용

안재형 · 김병용 · 김대훈 · 송재경 · 원항연*

국립농업과학원 농업미생물과

Application of Amplicon Pyrosequencing in Soil Microbial Ecology

Jae-Hyung Ahn, Byung-Yong Kim, Dae-Hoon Kim, Jaekyeong Song, and Hang-Yeon Weon*

Agricultural Microbiology Division, National Academy of Agricultural Science, Rural Development Administration, Suwon 441-853, Korea

Soil microbial communities are immensely diverse and complex with respect to species richness and community size. These communities play essential roles in agricultural soil because they are responsible for most of the nutrient cycles in the soil and influence the plant diversity and productivity. However, the majority of these microbes remain uncharacterized because of poor culturability. Next-generation sequencing techniques have revolutionized many areas of biology by providing cheaper and faster alternatives to Sanger sequencing. Among them, amplicon pyrosequencing is a powerful tool developed by 454 Life Sciences for assessing the diversity of complex microbial communities by sequencing PCR products or amplicons. This review summarizes the current opinions in amplicon sequencing of soil microbial communities, and provides practical guidance and advice on sequence quality control, aligning, clustering, OTU- and taxon-based analysis. The last section of this article includes a few representative studies conducted using amplicon pyrosequencing

Key words: Next generation sequencing, Amplicon pyrosequencing, Soil microbial ecology, Microbial diversity, Community structure

서 언

토양미생물은 토양 양분 순환에 관여하고, 식물 다양성과 생산성에 영향을 미치기 때문에 농업 시스템 중 가장 중요한 요인 중 하나이다. Van der Heijden et al. (2008)에 따르면 미생물은 토양의 탄소순환에 50%까지, 분해과정에는 100%까지 관여한다. 질소순환의 탈질 과정에는 60%까지, 인 순환의 경우에는 90%까지 미생물이 기여하고 있다. 따라서 토양미생물이 없다면 토양의 양분순환과 작물의 생육도 불가능하다.

토양은 지구상에서 종 다양성과 군집 크기의 측면에서 가장 크고 복잡한 미생물의 서식지로, 토양 1 g에는 10^9–10^{10}의 미생물이 살고 있고, 약 2,000–18,000 개의 유전체가 있는 것으로 추정된다 (Torsvik et al., 1996; Torsvik and Ovreas, 2002). 그러한 복잡한 다양성 때문에 토양미생물 군집은 표현형과 유전학적으로 특성을 파악하기가 매우 어렵다. 비록 분자생물학적 접근법으로 미생물 군집 연구가 되었지만, 토양의 미생물 다양성을 파악하는 것은 여전히 제한적이다. 더욱이 환경에 존재하는 미생물의 대부분 (>99%)은 실험실에서 배양이 되지 않기 때문에 (Amann et al., 1995), 그들의 유전 및 대사 특성, 생화학 기능, 생태학적 기능을 파악하는 것은 더욱 힘들다.

최근 454 FLX (Roche, USA), HiSeq (Illumina, USA), SOLiD (Life technologies, USA) 등의 대용량 염기서열 기법이 개발되었다. 이들 차세대 염기서열 분석 기술은 비용이 저렴하고 속도가 빠르기 때문에 Sanger sequencing의 대체물로서 생물학의 여러 분야에 혁명과도 같은 영향을 주고 있다. 본 논문에서는 차세대 염기서열 분석 기술 중 증폭 파이로시퀀싱에 초점을 맞추어, 미생물의 다양성 분석 원리와 실제를 고찰하고자 한다.

본 론

최근 미생물 생태 분야에서는 전통적인 배양법에 의존하지 않고, 16S rRNA 유전자 분석 등의 분자생물학적 방법이 일반적으로 사용되고 있다. 세균 군집 구조의 비배양적 방법은 모든 세균 군집 DNA의 분리, 16S rRNA 유전자의 증폭, 대장균 숙주 내로 클로닝 (cloning), 클로닝된 증폭 유

접수 : 2012. 11. 5 수리 : 2012. 11. 23
*연락저자 : Phone: +82312908478
E-mail: why@korea.kr

Table. 1. List of NGS sequencing platforms and their expected throughputs, error types and error rates. Each platform has distinct advantages owing to cost, error rate, read length, and so on (adapted from Scholz et al., 2012).

Platform	Run time	Read length	Throughput per run	Error type	Error rate
	h	bp	Mb		%
Roche					
454 FLX+	18-20	700	900	Indel	1
454 FLX Titanium	10	400	500	Indel	1
454 GS	10	400	50	Indel	1
Illumina					
GAIIx	14	2 × 150	96,000	Substitution	>0.1
HiSeq 2000	8	2 × 100	400,000	Substitution	>0.1
HiSeq 2000 V3	10	2 × 150	<600,000	Substitution	>0.1
MiSeq	1	2 × 150	1000	Substitution	>0.1
Life technologies					
SOLiD 4	12	50 × 35	71,000	A-T Bias	>0.06
SOLiD 5500xl	8	75 × 35 PE 60 × 60 MP	155,000	A-T Bias	>0.01
Ion torrent					
PGM 314 Chip	3	100	10	Indel	1
PGM 316 Chip	3	100+	100	Indel	1
PGM 318 Chip	3	200	1000	Indel	1
Pacific biosciences					
RS	14/8 Smart Cells	1500	45/SC	Insertions	15

전자의 염기서열 분석을 수반한다. 그러나 이와 같은 연구 방법은 시간이 많이 소요되고, 처리할 수 있는 시료수의 제한이 있을 수 밖에 없다. 최근 이 방법은 클로닝 과정이 필요 없는 차세대 염기서열 분석 기술로 빠르게 대체되고 있다.

차세대 염기서열 분석 기술은 가동시간, 염기서열 길이, 처리량에 따라 다양한 기종이 출시되어 있다 (Table 1). Roche 454는 염기서열의 해독 길이가 길고, Illumina 기종은 한 번 가동 시 대량의 염기서열을 얻을 수 있다는 장점이 있다. 이들 방법들은 유전체학, 메타유전체학, 메타전사체학, 증폭 시퀀싱 등의 다양한 생명과학 분야에 장단점을 살려 널리 사용되고 있다. 특히 Roche 454의 파이로시퀀싱은 분석된 염기서열이 길기 때문에 미생물생태 분야의 증폭 시퀀싱 분야에서 독보적인 위치를 차지하고 있다. 반면 Illumina HiSeq 2000의 경우는 파이로시퀀싱보다 많은 염기서열수를 생성하지만 염기서열 길이가 짧아 미생물 분류를 위한 정보량이 상대적으로 부족하다는 단점이 있다. 그러나 향후 해독되는 염기서열 길이의 길어진다면 군집분석을 위한 더 강력한 도구가 될 가능성을 배제할 수 없다. Caporaso et al. (2012)은 이 플래폼의 가능성을 제시하고 있으며, 이미 여러 연구자에 의해 이용된 바 있다 (Bartram et al., 2011; Knight et al., 2011; Zhou et al., 2011).

1. 파이로시퀀싱을 이용한 다양성 분석

가. 파이로시퀀싱의 원리

파이로시퀀싱은 DNA 중합효소가 뉴클레오티드 한 개를 합성할 때 생성되는 pyrophosphate를 검출하여 염기서열을 해독하는 기술이다 (Fig. 1). 즉, pyrophosphate가 방출되면서 sulfurylase에 의해 APS (adenosine-5'-phosphosulfate)가 ATP (adenosine triphophate)로 전환되고, 생성된 ATP는 luciferase에 의해 빛을 방출한다. 이 때 빛의 양은 삽입된 뉴크레오티드의 수에 비례한다. GS-FLX platform은 clonal amplification을 위해 emulsion PCR을 이용하며, 이 과정을 통해 bead에서 증폭된 DNA 단편이 loading된다. 이때 하나의 reactor에 한 개의 bead만 들어가도록 reactor의 반경이 최적화 되어 있기 때문에 하나의 reactor에서 나온 신호는 하나의 DNA 단편에서 유래한 신호라고 판단할 수 있다.

파이로시퀀싱은 소위 바코드 전략, 즉 PCR 프라이머에 독특한 염기서열 (바코드)을 삽입함으로써, 시퀀싱 반응을 할 때, 동시에 많은 시료를 혼합하여 분석할 수 있다. 파이로시퀀싱은 유전체 DNA 또는 증폭산물을 이용하는 방법이

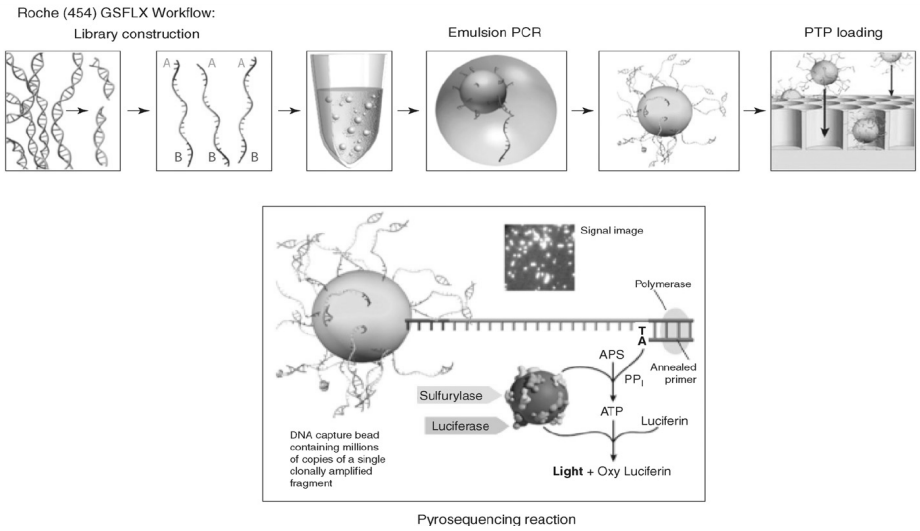

Fig. 1. 454 Workflow: library construction ligates 454-specific adapters to DNA fragments and couples amplification beads with DNA in an emulsion PCR to amplify fragments before sequencing. The beads are loaded into the picotiter plate (PTP). The bottom panel illustrates the pyrosequencing reaction that occurs on nucleotide incorporation to report sequencing by synthesis (adapted from Mardis, 2008).

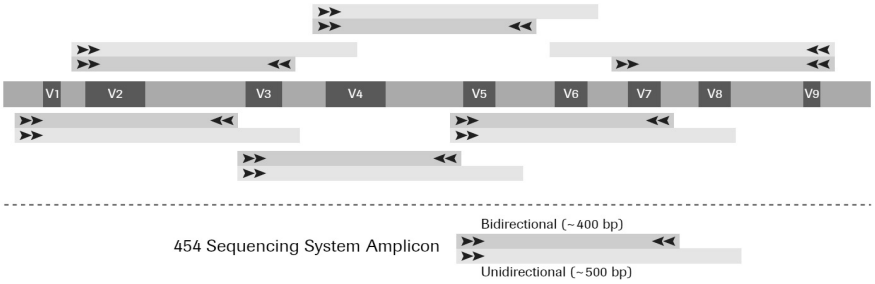

Fig. 2. Illustration of the 16S rRNA subunit showing possible candidate variable region coverage (adapted from www.roche.com).

있다. 전자의 경우는 Fig. 1의 초기단계에 DNA의 단편을 만들고 적당한 어댑터를 부착시킨 후 파이로시퀀싱을 하는 방법이다. 반면 후자는 PCR 증폭산물 (증폭체)을 이용하기 때문에 증폭 시퀀싱이라 정의할 수 있다 (편의상 이후부터는 파이로시퀀싱이란 용어를 사용하기로 하겠다). 즉 Fig. 2 (16S rRNA 유전자를 이용한 예)와 같이 바코드가 부착된 적당한 프라이머를 이용하여 마커유전자를 증폭한 후 단방향 또는 양방향으로 파이로시퀀싱을 수행하는 방법이다.

나. 다양성 분석에 사용되는 유전자

세균과 고세균 다양성 분석에는 리보솜 RNA (ribosomal RNA, rRNA)의 소 단위체 (16S rRNA) 영역의 염기서열이 이용된다. 16S rRNA 유전자는 다른 유전자에 비해 다양성 분석에 있어 많은 이점을 갖고 있다. 1) 변이가 적은 (보존이 잘된) 영역은 PCR 프라이머 디자인에 사용되며 2) 변이가 많은 영역을 통해서 군집 구성원의 정확한 분류학적, 계통학적 동정이 가능하며 3) 분류군 사이의 수평적 유전자 전이 (horizontal gene transfer)가 드물며 4) 광범위한 16S rRNA 염기서열 데이터가 데이터베이스로 축적되어 있어, 정확한 분류가 가능하다. 진균의 경우는 18S rRNA 유전자나 ITS 영역 (internal transcribed spacer region)이 이용된다.

다. 파이로시퀀스 분석 절차

파이로시퀀싱을 통해 한 시료 당 얻을 수 있는 염기서열의 수는 몇 개의 시료를 한 번의 분석에 투입하느냐에 따라 달라지지만, 보통 수 천 개에서 수 만개에 이른다. 따라서 컴퓨터의 도움 없이는 분석이 불가능하다. 파이로시퀀싱 데이터의 해석을 위해 다양한 종류의 프로그램 및 알고리즘이 개발되었으며, 이에 따라 문헌에 나타난 파이로시퀀싱 데이터의 분석 방법도 다양하다. Fig. 3에 파이로시퀀싱으로 얻은 데이터를 이용하여 미생물의 군집 구조를 분석할 때의 일반적인 분석 경로 (pipeline)를 도식화하였다.

파이로시퀀싱을 통해 얻어진 염기서열은 먼저 스크린 과정 및 키메라 염기서열 제거 과정을 통해 PCR 및 파이로시퀀싱 과정의 오류로 생성된 염기서열 (artifact)을 제거해야 한다. 이들로 인해 실제 미생물 다양성이 과대 평가될 수 있기 때문이다. 다음으로 염기서열을 정렬 (alignment)하는 과정을 수행한다. 공통 조상 (common ancestor)에서 유래

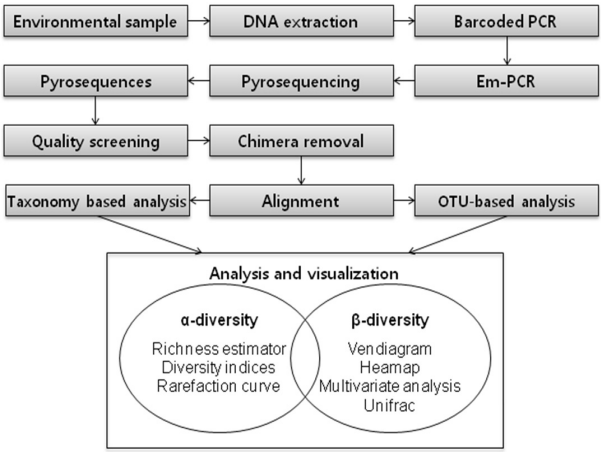

Fig. 3. General scheme of the amplicon pyrosequencing analysis of microbial community.

한 유전자는 오랜 시간 진화를 겪으면서 염기의 종류가 변할 뿐만 아니라 (substitution), 새로운 염기가 삽입 (insertion)되거나 삭제 (deletion)되면서 그 길이가 변한다. 따라서 염기서열 간의 유사도 비교를 위해서는 염기서열 내의 각각의 염기가 그 본래의 위치 (homologous position)에 오도록 모든 염기서열들을 정렬해야 한다. 이는 실제로 각 염기서열 내에 틈 (gap)을 삽입함으로써 이루어진다.

다음으로 시료 안에 미생물 종류와 그 비율을 알기 위해 각각의 염기서열들을 특정한 기준에 따라 분류하는데, 이 과정은 크게 두 가지가 있다 (Schloss and Westcott, 2011). 첫째는 기존의 미생물 분류학에 따라 명명된 염기서열과의 유사도에 근거하여 시료에서 얻은 염기서열을 동정하는 것이다 (미생물분류학 기반 군집구조 분석). 둘째는 시료에서 얻은 염기서열 간의 유사도에 근거하여 염기서열들을 묶고, 이를 한 단위로 하여 시료 내의 미생물 다양성을 평가하는 것이다. 이러한 단위를 조작상분류단위 (Operational taxonomic unit, 이하 OTU)라고 하며, 이때 사용하는 유사도는 연구자가 임의로 정의할 수 있다 (OTU 기반 군집구조 분석).

마지막 단계는 위에서 얻은 결과를 바탕으로 시료 내의 미생물 다양성을 평가하고 서로 다른 시료의 미생물 군집구조를 비교한다.

1) 스크리닝 (선별과정)

스크리닝 단계에서는 본래 시료 내에 존재하지 않았지만 PCR 과정과 파이로시퀀싱과정에서 발생한 오류 염기서열 (artifact)을 제거한다. PCR 과정에서 발생하는 오류는 사용한 중합효소의 종류 및 PCR 사이클 수에 영향을 받는다 (Acinas et al., 2005). 파이로시퀀싱 역시 반응 고유의 성질에 의해 염기서열을 읽는 과정에서 오류가 발생하며 그 비율은 0.5–1.1%인 것으로 알려져 있다 (Gilles et al., 2011; Huse et al., 2007). Huse et al. (2007)은 GS20 시스템을 이용한 분석 결과에 근거하여 파이로시퀀싱 과정에서 생기는 오류를 0.5% 이하로 낮추기 위한 방법으로서, 1) 정확히 결정되지 않은 염기 (N)를 하나라도 가진 염기서열의 제거, 2) 길이가 평균으로부터 많이 벗어난 염기서열의 제거, 3) 프라이머 부분의 염기서열에 오류가 있는 염기서열의 제거를 제시했다. 이 기준은 여러 연구에서 표준 스크리닝 방법으로 사용되었으며, 그 외에 파이로시퀀싱은 반응의 특성상 같은 염기가 연속된 경우 (homopolymer), 염기서열을 잘못 읽는 경우가 많은데 (Gilles et al., 2011; Huse et al., 2007), 16S rRNA 유전자의 경우 homopolymer의 길이는 일반적으로 8개 이하이므로 9개 이상의 homopolymer를 가진 염기서열을 제거하는 것이 좋다.

2) 키메라 제거

키메라는 PCR 과정에서 두 종 이상의 염기서열이 합쳐져서 새로운 염기서열이 생성되는 경우에 발생한다. 이는 프라이머가 미생물 A의 DNA와 결합한 후 중합효소에 의해 염기를 연장시키는 과정이 중단되었을 때 발생한다. 즉, 불완전하게 연장된 DNA가 새로운 프라이머로 작용하여 미생물 B에서 유래한 DNA에 결합하고, 그 뒤로 염기가 신장되면 A와 B의 염기서열이 혼합된 새로운 DNA가 생성된다 (von Wintzingerode et al., 1997). 파이로시퀀싱으로 얻은 염기서열의 길이가 200 bp 이하였던 초기에는 키메라가 거의 발생하지 않는 것으로 생각되었으나 (Huse et al., 2008), 현재 ~500 bp까지 분석이 가능한 버전 (454 FLX Titanium)이 출시되면서 키메라 제거 과정이 미생물 군집 분석에 있어서 중요해졌다. 키메라는 다른 파이로시퀀싱 에러와 같이 미생물의 종 다양성을 증가시키고, 실제로 존재하지 않는 미생물을 군집분석에 포함시키기 때문에 반드시 제거되어야 한다. 키메라 스크리닝은 일반적으로 16S rRNA 유전자를 두 부분으로 나눈 후 각 부분에 대하여 유사도가 가장 높은 염기서열을 16S rRNA 유전자 데이터베이스 상에서, 또는 같은 시료 안에 있는 다른 16S rRNA 유전자에 대하여 찾고, 이 두 염기서열 간 유사도를 비교하여 키메라 여부를 판정한다. 즉 두 염기서열의 유사도가 낮을 경우 본래의 염기서열은 키메라 확률이 높다고 판정할 수 있다. 키메라 염기서열 제거용 프로그램으로서 Bellerophon (Huber et al., 2004), Pintail (Ashelford et al., 2005), Chimera slayer (Haas et al., 2011), Perseus (Quince et al., 2011), UCHIME (Edgar et al., 2011), DECIPHER (Wright et al., 2012) 등 다양한 프로그램들이 개발되었으나, 키메라의 완벽한 제거는 어렵다.

3) 정렬 (alignment)

염기서열을 정렬하는 방법에는 pairwise alignment, multiple alignment, profile-based alignment가 있다 (Schloss, 2009). Pairwise alignment는 두 개의 염기서열을 정렬하는 방법이고, multiple alignment는 3개 이상의 염기서열을 정렬하는

방법이다. Pairwise aligner로는 BLAST (Altschul et al., 1990)와 ESPRIT (Sun et al., 2009) 등이 있으며, multiple aligner로는 ClustalW (Chenna et al., 2003), MUSCLE (Edgar, 2004) 등이 있다. 다수의 염기서열 간에 의미 있는 유사도 분석을 위해서는 multiple alignment 과정이 필수적이지만 (Schloss, 2009), 이 방법은 컴퓨터에 대한 부하가 높고 16S rRNA의 2차 구조를 고려하지 않는다는 단점이 있다. 16S rRNA 염기서열의 정렬에 있어 RNA 2차 구조의 고려는 중요하다 (Peplies et al., 2008). Profile-based aligner는 rRNA의 2차 구조를 고려한 정렬 방법으로서 RDP aligner (Cole et al., 2009), ARB fast aligner (Ludwig et al., 2004), SINA aligner (Pruesse et al., 2012), Greengenes aligner (http://greengenes.lbl.gov) 등이 있다.

4) 미생물 군집구조 분석

가) 미생물 분류학 (taxonomy) 기반

데이터베이스의 분류된 염기서열을 이용하여 동정한 후 군집구조를 분석하는 방법이다. 이 방법은 OTU에 근거한 방법에 비해 컴퓨터에 대한 부하가 적다는 장점이 있다. 반면 순수분리가 안되었거나 어려운 미생물의 경우 연구자 또는 데이터베이스 간 분류 체계의 차이가 있을 수 있다는 단점이 있다 (DeSantis et al., 2006). 아래에 온라인 상에서 이용 가능한 데이터베이스들을 간략히 소개하였다.

RDP 미국 미시건 주립대학에서 관리하는 데이터베이스로 2012년 12월 현재 세균 2,518,232개, 고세균 120,405개 및 미분류된 520개의 염기서열을 포함하여 총 2,639,157개의 16S rRNA 염기서열을 보유하고 있다. 이 중 97,919개는 표준균주로부터 유래한 것이며 2,274,924개는 비배양 (uncultured) 미생물로부터 얻어진 것이다. 데이터베이스는 DDBJ, EMBL, GenBank로부터 매달 업데이트된다. 미생물은 Garrity et al. (2007)의 체계에 따라 분류되고, *Acidobacteria*, *Verrucomicrobia*, OP11 등은 개별 연구자들의 연구결과들을 기반으로 하고 있다 (http://rdp.cme.msu.edu/).

GenBank 미국 국립보건원의 NCBI (National Center for Biotechnology Information)에서 관리하는 데이터베이스로 유럽의 EMBL, 일본의 DDBJ와 매일 데이터 교환을 한다. 또한 이들과 협력하여 염기서열에 기초한 생물분류법을 수립하여 각 염기서열에 분류명을 부여하였다. 380,000 종 이상의 생물 (세균, 고세균, 진핵생물, 바이러스 포함)에서 유래한 염기서열을 포함하고 있으며 매달 3,800개 이상의 새로운 분류군이 추가된다. 다른 데이터베이스와는 달리 GenBank에서는 염기서열에 대한 품질 검사와 염기서열의 정렬이 수행되지 않는다. BLAST 프로그램들을 이용하여 GenBank 상에서 유사한 염기서열을 검색할 수 있다 (Benson et al., 2011) (http:// www.ncbi.nlm.nih.gov/).

Greengenes 미국 버클리 대학의 Lawrence Berkeley National Lab에서 운영하고 있고, 2012년 12월 현재 세균 및 고세균의 16S rRNA 염기서열 1,049,116개를 포함하고 있다. 모든 염기서열은 NAST aligner를 이용하여 정렬되어 있고, Bellerophon 프로그램을 이용하여 키메라 검사가 수행되며 정해진 기준에 따라 키메라 여부를 표시하고 원본으로 추정되는 염기서열 및 그 유사도가 표시된다. 여섯 곳의 관리자 (Pace, Hugenholtz, Ludwig, NCBI, G2_Chip, RDP)에 의해 제시된 분류체계가 모두 표시되며, 이들 간 비교도 가능하다. Greengenes 데이터베이스에서는 분류의 단위로서 속 (genus) 아래에 OTU가 위치하며 (DeSantis et al., 2006), OTU 아래에는 순수분리된 미생물 또는 비배양된 미생물의 16S rRNA 유전자가 위치한다 (http://greengenes.lbl.gov/).

Silva 독일 막스프랑크 해양미생물연구소의 미생물 유전체 그룹이 관리하는 데이터베이스로 세균, 고세균, 진핵생물을 모두 포함하며, small subunit rRNA 뿐만 아니라 large subunit rRNA도 포함하고 있다. 모든 염기서열은 품질 확인 과정을 거치며, Pintail 프로그램을 이용하여 키메라를 검사한다. 2차 구조를 고려하여 수작업으로 정렬된 58,795개의 염기서열을 참조 염기서열 (reference sequences)로 하여 Silva INcremental Aligner (SINA) 프로그램을 이용하여 모든 염기서열이 정렬되어 있다 (Pruesse et al., 2012). Silva 데이터베이스는 Parc, Ref, NR의 세 개의 데이터베이스를 제공하는데, Parc 데이터베이스는 300 bp 이상의 염기서열을 포함하며 미생물다양성 분석에 적합하다. Ref 데이터베이스는 Parc 데이터베이스에서 특정 길이 이하의 염기서열 및 정렬 품질이 낮은 염기서열을 제외한 데이터베이스로서 계통분류학적 분석 및 probe 디자인에 적합하다. NR 데이터베이스는 Ref 데이터베이스의 염기서열들을 98% 유사도를 기준으로 묶고 각 그룹에 대하여 한 개의 대표 염기서열만을 추출하여 만들어 진다. 2012년 11월 현재 각각의 데이터베이스는 3,194,7782개, 739,633개, 286,858개의 small subunit rRNA 염기서열을 포함하고 있다. 미생물 분류는 ENA-EBI (EMBL)의 분류법을 기반으로 하며, Greengenes와 RDP의 미생물 분류체계도 표시되어 있다. 염기서열 데이터베이스 관리 및 분석 프로그램인 ARB (Ludwig et al., 2004)에서 사용 가능한 형식을 제공되며 4-8개월 단위로 업데이트된다 (http://www.arb-silva.de/).

EzTaxon-e 한국의 천연구소에서 제공하는 데이터베이스로 유효 출판되었거나 출판 중인 표준 균주와 Candidatus 상태인 염기서열을 포함하며, 비배양 미생물의 경우 3% 유사도를 기준으로 하여 그 대표 염기서열을 포함하고 있다.

세균, 고세균, 진핵생물 모두 망라하며 총 37,333개의 염기서열이 데이터베이스로 구축되어 있다. EzTaxon-e에서는 모든 염기서열이 종 (species)부터 문 (phylum)까지 분류되어 있으며 분류명이 없는 경우, 임시로 분류명이 부여된다. 예를 들어 과 (family) 명이 알려지지 않은 *Xenohaliotis* 속은 *Xenohaliotis_f* 과로 표기하며, 비배양 미생물에서 유래한 경우 그 염기서열의 GenBank 등록 번호 (accession number)를 이용하여 나타낸다. 예를 들어 AB177171_g는 등록 번호 AB177171를 기반으로 하는 속을 가리킨다 (http://eztaxon-e.ezbiocloud.net/).

파이로시퀀싱으로부터 얻은 염기서열을 데이터베이스의 염기서열과 비교하고 동정을 하기 위한 알고리즘을 몇 가지 소개하면 다음과 같다.

Global Alignment for Sequence Taxonomy (GAST) GAST (Huse et al., 2008)에서는 rRNA 데이터베이스 상에 있는 16S rRNA 유전자 염기서열에서 파이로시퀀싱에 사용된 영역 (예, V3)만을 추출해서 새로운 데이터베이스를 만든다. BLAST를 이용하여 분석하고자 하는 염기서열과 유사도가 높은 100개의 염기서열을 새로운 데이터베이스 상에서 추출하고 MUSCLE 프로그램을 이용하여 multiple alignment를 수행한 후 염기서열 간의 거리를 계산한다. 이 때 거리가 가장 가까운 염기서열의 분류명을 분석하고자 하는 염기서열의 분류명으로 부여한다. 거리가 동일한 염기서열이 둘 이상 있을 경우 일치하는 분류 수준까지만 분류명을 부여한다. 즉 두 염기서열이 과 수준까지만 일치한다면 그 과까지만 분류명을 부여한다.

Naïve Bayesian classifier Naïve Bayesian classifier (Wang et al., 2007)는 RDP의 미생물 분류 프로그램에서 사용하는 방법이다. 먼저 알려진 표준균주의 16S rRNA 염기서열에 대하여 8개의 염기로 정렬 가능한 64,000여 개의 염기서열 (단어, word)이 나타나는 빈도를 계산한다. 다음으로 분류하고자 하는 염기서열에 나타나는 8개의 염기로 구성된 염기서열 중 일부를 취하여 이 염기서열이 각각의 표준균주에서 발견될 확률을 구하고, 확률이 가장 높은 표준균주의 속 명을 부여한다. 이 작업은 100회 수행되며, 이 중 그 표준균주로 할당된 횟수를 계산하여 그 신뢰값을 %로 나타낸다. 속 이상의 분류군에 대해서는 각 분류군에 포함된 모든 속에 대한 결과값을 합산한다. 이 방법은 염기서열의 정렬이 필요하지 않다는 장점이 있고, 속까지만 분류를 수행한다.

CLcommunity 동정법 미생물 군집분석 프로그램인 CLcommunity에서 사용하는 방법으로 EzTaxon-e 데이터베이스의 염기서열과의 유사도를 비교하여 다음 규칙에 의해 분류명을 부여한다.

유사도가 97% 이상일 경우 같은 종 (species)
유사도가 97% 미만 94% 이상일 경우 같은 속 (genus)
유사도가 94% 미만 90% 이상일 경우 같은 과 (family)
유사도가 90% 미만 85% 이상일 경우 같은 목 (order)
유사도가 85% 미만 80% 이상일 경우 같은 강 (class)
유사도가 80% 미만 75% 이상일 경우 같은 문 (phylum)

염기서열이 특정 종으로 동정되지 않을 경우 해당되는 분류군에 _uc를 붙여 새로운 분류군을 만든다. 예를 들어 가장 높은 유사도를 가지는 염기서열이 *Nocardia asteroides* 이고 그 유사도가 92%일 경우 이 염기서열은 *Nocardiaceae* _uc라는 새로운 속으로 동정된다 (http://www.chunlab.com/i).

나) 조작상분류단위 (Operational taxonomic units, OTU) 기반

OTU에 근거하여 미생물의 군집구조를 분석하는 과정은 1) 각 염기서열간의 거리를 계산하는 과정과 2) 염기서열들을 OTU로 할당하는 과정으로 나눌 수 있다. 각 염기서열간의 거리는 염기서열을 정렬한 후 계산된다. 거리 계산에 있어 정렬된 염기서열에서 연속된 틈 (gap)을 어떻게 처리할 것인가를 고려해야 한다. Schloss (2010)는 연속된 틈을 하나의 틈으로 간주할 것을 추천하고 있다.

	1	2	3	4		5		6	7	8	9	10
Sequence A :	A	T	G	C	A	T	G	C	A	T	G	C
Sequence B :	A	C	G	C	-	-	-	C	A	T	C	C

위의 예에서 A, B 두 염기서열간의 거리는 3/10=0.30으로 계산된다. 즉 2번과 9번 위치에 일치하지 않는 염기가 있으며, 5번 위치의 3개의 연속된 틈은 1개의 불일치로 간주하여, 총 10개의 염기 중 3개가 불일치하는 하는 것으로 계산한다. 다수의 염기서열에 대한 거리 계산 결과는 아래와 같이 거리 행렬로 나타낼 수 있다. 여기서 A, B, C, D, E는 서로 다른 염기서열을 의미하고 내부의 숫자는 각 염기서열간의 거리를 나타낸다.

	A	B	C	D	E
A	0.00				
B	0.02	0.00			
C	0.04	0.06	0.00		
D	0.09	0.10	0.11	0.00	
E	0.11	0.10	0.09	0.01	0.00

다음으로 이 계산결과를 이용하여 각 염기서열들을 OTU로 묶는다. 이 때 1) 염기서열들을 묶어 나가는 방식 (클러스터링 알고리즘)과 2) OTU간의 최소 거리를 얼마로 할 지

를 고려해야 한다.

클러스터링은 거리가 가장 가까운 두 염기서열부터 시작하여 클러스터링 알고리즘에서 정의한 거리를 이용하여 염기서열들을 하나의 클러스터로 묶어 나가는 과정이다. 일반적으로 사용되는 클러스터링 알고리즘에는 1) nearest neighbor, 2) furthest neighbor, 3) average neighbor가 있다. Nearest neighbor에서 두 클러스터간의 거리는 각 클러스터 안에 있는 염기 서열들간의 가능한 모든 거리 중 가장 가까운 거리로 정의한다 (Fig. 4a). Nearest neighbor 방법을 사용하고 OTU 간의 최소 거리를 3%로 정한다면, 한 OTU에 포함된 각 염기서열들은 거리가 3% 이하인 염기서열을 최소 한 개 이상 같은 OTU에서 찾을 수 있다. Furthest neighbor에서 두 클러스터간의 거리는 각 클러스터 안에 있는 염기 서열들간의 가능한 모든 거리 중 가장 먼 거리로 정의한다 (Fig. 4b). Furthest neighbor 방법을 사용하고 OTU 간의 최소 거리를 3%로 정한다면, 한 OTU에 포함된 각 염기서열들 간의 거리는 모두 3% 이하이다. Average neighbor에서 두 클러스터간의 거리는 각 클러스터 안에 있는 염기 서열들간의 가능한 모든 거리의 평균으로 정의한다 (Fig. 4c).

초기에는 furthest neighbor를 기본 방법으로 사용하였으나 (Schloss and Handelsman, 2005), 최근 연구 결과 이 방법이 실제 미생물의 다양성을 과대 평가하는 것으로 나타났으며 (Huse et al., 2010), 이에 대한 대안으로서 average neighbor (Schloss and Westcott, 2011) 또는 nearest neighbor 적용 후 average neighbor (Huse et al., 2010) 방법을 적용

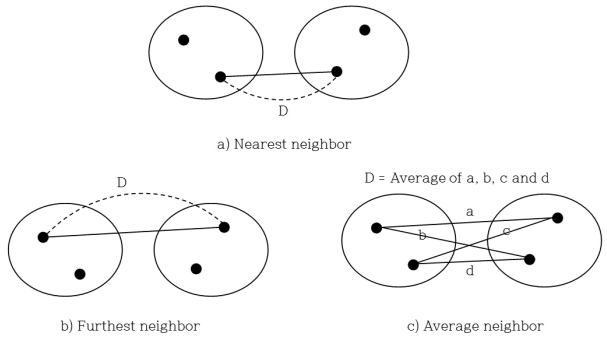

Fig. 4. Three clustering algorithms. D represents the distance between two clusters.

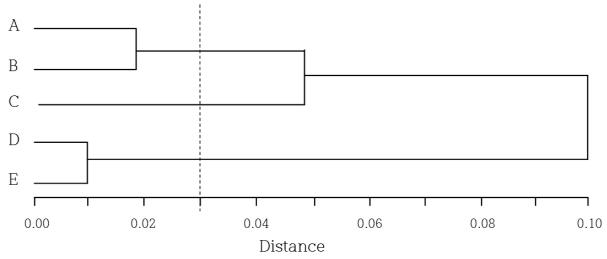

Fig. 5. A dendrogram obtained from the distance matrix in the example by using average neighbor method.

하는 것이 제안되었다.

세균 군집분석의 경우 OTU간의 최소 거리는 일반적으로 3%가 권장된다. 이는 1) 16S rRNA 유전자를 이용하여 종을 구분 짓는 경계선인 97% 유사도 (Stackebrandt and Goebel, 1994), 2) 한 염색체 내 rRNA operon의 변이 정도, 3) 파이로시퀀싱 과정에서 발생하는 염기서열의 인위적 변이들을 고려한 결과이다 (Kunin et al., 2010).

Fig. 5는 앞서 언급한 거리 행렬에 average neighbor 알고리즘을 적용하여 클러스터링을 수행한 결과이다. 여기서 아래에 표시한 숫자는 각 염기서열 간, 또는 클러스터 간의 거리를 의미한다. OTU간의 최소 거리를 3%로 정의한다면 A와 B, D와 E가 각각 하나의 OTU가 되며 C는 단독으로 하나의 OTU가 된다. 결국 다섯 개의 염기서열은 총 3개의 OTU로 분류된다.

염기서열 간 유사도를 계산하고 클러스터링을 수행하는 프로그램으로서 Cd-hit가 있다 (Li and Godzik, 2006). Cd-hit에서는 염기서열을 정렬하지 않는다. 대신 공유되는 특정 염기서열의 갯수를 조사하여 염기서열 간 유사도를 구하고 클러스터링을 수행한다. 이 프로그램은 계산과정이 빠르고 컴퓨터에 대한 부하가 적어 대량의 염기서열을 분석하는데 사용될 수 있다.

2. 파이로시퀀스 분석 시 고려 사항

토양미생물 생태 연구에 있어 의미 있는 데이터를 도출하기 위해서는 시료채취 시간, 위치, 계절, 시료량, 혼합, 복합시료 조제, 반복 등을 고려해야 한다. 보통 DNA는 소량의 시료 (0.1-5g 정도)에서 추출되기 때문에 추출 전 균질화 및 충분한 반복을 두어 시료가 대표성을 가지도록 해야 한다. 또한 시료는 현장과 실험실에서 분석되는 시간 사이에 변화가 없어야 하고 의도하지 않게 다른 시료와 섞인다거나 조제, 이송, 분석 도중에 다른 시료와 접촉으로 오염되지 않아야 한다.

가. 시료의 보존

미생물 DNA는 시료 채취 즉시 추출하는 것이 원칙이며, 이것이 가능하지 않다면 시료는 동결시켜야 된다 (Rochelle et al., 1994). 토양 세균 군집 구조는 보존 조건에 큰 영향을 받지 않는다는 보고도 있으나 (Dolfing et al., 2004; Klammer et al., 2005) Tzeneva et al. (2009)은 저장조건에 따라 토양시료는 미생물 군집 다양성과 조성이 크게 영향을 받는다고 하였다.

Lauber et al. (2010)은 시간과 온도를 달리하여 보존한 토양, 인체피부, 인체 분 시료를 대상으로 파이로시퀀싱 방법을 이용하여 계통 및 분류학적 군집분석을 수행하였다. 그 결과, 시료 출처와 종류가 DNA 추출 전 보존방법 (온도, 기간)에 비해 미생물 군집구조에 더 큰 영향을 줄 수 있다고

보고하였다.

Wu et al. (2010a)은 분석 전 시료 보존 방법과 DNA 추출 방법의 영향을 조사한 결과, i) 보존방법에 따른 군집조성 변이에 대한 영향은 시료 간 변이에 비해 작기 때문에, 보존 방법은 시험 성격에 맞게 선택될 수 있고 ii) DNA 추출 방법은 군집구조 해석에 큰 영향을 줄 수 있다는 결론에 도달하였다.

나. DNA 추출

토양 군집의 분자생물학적 분석을 저해할 수 있는 중요한 화학적 인자 중에는 양이온 또는 음이온 교환능을 지닌 부식산, 점토, 유기물 등이 있다. 토양 점토 입자와 유기물은 음전하를 띠고 있으며, 음전하를 띤 점토 입자는 수화된 양이온에 둘러싸이게 되어 양전하의 편재화된 영역을 만들게 된다. 여기에 일반적인 토양 pH에서 알짜 음전하를 띤 미생물이 부착된다. 세균이 고체 표면에 부착하게 되면, 토양 매질에서 세포 유래의 DNA/RNA와 세포를 분리하기 어렵다. 토양 유형은 DNA/RNA 추출 효율에 큰 영향을 끼친다. 즉 점토, 유기물, 부식산 등이 많을 경우 핵산 추출에 문제가 되며, 특히 PCR을 수행할 때 *Taq* DNA 중합효소를 저해할 수 있다. 따라서 부식산은 DNA를 정량하거나 증폭하기 전에 제거되어야 한다.

환경시료에서 DNA를 추출하기 위해서는 물리적, 화학적, 효소학적 방법 등을 사용하며, 이 세 가지 방식이 단독 또는 조합하여 사용되고 있다. 그 외에 상용화되어 있는 것들도 위의 방법을 변형하거나 조합한 방법이다 (Yoo et al., 2009). 한편 Ranjard et al. (2003)은 토양 시료량이 분자미생물생태 분석에 미치는 영향을 평가하여 유전적 군집구조의 대표성를 가지기 위해서는 최소 1 g의 토양에서 DNA를 추출할 것을 권하고 있다.

다. 증폭 조건

증폭 파이로시퀀싱은 특정 유전자만을 증폭하기 위해 PCR을 사용한다. 즉 파이로시퀀싱을 통해 분석되는 염기서열은 본래 시료에 존재했던 16S rRNA 유전자가 아니라 PCR을 통해 새롭게 생성된 염기서열이다. 따라서 분석된 염기서열은 본래의 것과 다를 수 있으며, 이는 크게 두 가지 원인에 기인한다. 첫째는 키메라 염기서열의 생성이나 DNA 중합효소의 오류로 인해 염기서열이 변하는 경우로서, 이는 앞 절에서 설명하였다. 둘째는 서로 다른 염기서열 간의 상대적인 풍부도가 PCR 과정 동안 변하는 경우이다.

서로 다른 염기서열 간의 상대적인 풍부도가 PCR 후 변하는 것은 프라이머와의 결합력 차이 (Hongoh et al., 2003; Polz and Cavanaugh, 1998)와 PCR 과정에서의 포화정도의 차이에 기인한다 (Suzuki and Giovannoni, 1996). 즉, 전자는 특정 프라이머 염기서열이 미생물 A와는 일치하지만 B와는 일치하지 않기 때문에 미생물 A가 선택적으로 증폭되기 때문에 발생한다. 반면 후자는 PCR이 대수단계에서 포화단계로 진행됨에 따라 초기 농도가 높은 염기서열은 re-annealing 빈도가 높아 증폭속도가 감소하는 반면, 농도가 낮은 염기서열은 계속 증폭이 되어 두 종 간 풍부도의 차이가 감소되기 때문에 발생한다.

이러한 문제점들은 적절한 프라이머를 선택하거나 (Watanabe et al., 2001), annealing 온도를 낮추거나 PCR 반복 횟수를 감소시켜 해결될 수 있다 (Acinas et al., 2005; Ishii and Fukui, 2001). 특히 PCR 반복 횟수의 감소는 PCR 과정에서 발생하는 다양한 변이를 감소시키는데 중요하다 (Qiu et al., 2001; Suzuki and Giovannoni, 1996; Wu et al., 2010b). 그러나, 아직까지 표준화된 프로토콜은 없으며, 연구자가 각자의 실험목적, 신뢰성, 비용, 효율성을 판단하여 적절한 PCR 조건을 선택해야 할 것이다.

라. 분석 경로

파이로시퀀싱을 통해 생산된 데이터의 분석은 한 종류의 응용프로그램만으로 수행될 수 없으며, 앞 절에서 제시된 분석경로의 각 단계별로 여러 연구자들이 개발한 응용프로그램들을 선택하여 이용해야 한다. Mothur 프로그램 (Schloss et al., 2009)은 현재 가장 많은 명령어와 옵션을 포함하고 있기 때문에 유연성이 가장 높다고 할 수 있다. 그러나 명령어 방식이라 사용자 편의성이 낮고, 그래프 등의 시각화는 별도의 작업이 필요하다. 또한 특정 명령어는 높은 사양의 컴퓨터 장비를 이용해야만 실행이 가능하다. RDP (rdp.cme.msu.edu)는 온라인상에서 파이로시퀀싱 데이터를 분석할 수 있는 Pyrosequencing Pipeline을 제공한다. 염기서열 정렬, OTU 생성, 다양성 지수 계산, 미생물분류 등 기본적인 분석을 수행할 수 있어 초보자에게 가장 적합할 것으로 판단된다. RDP에서 생성된 결과 파일은 SPADE, EstimateS, R, mothur 등의 다른 프로그램에서도 사용이 가능하다. QIIME (Caporaso et al., 2010)은 기본적인 분석도구 외에도 출판을 위한 고화질의 그래프까지 출력할 수 있고 분석속도도 빠르다. 천연구소에서 제공하는 CLcommunity™ 프로그램은 사용자가 직접 분석을 수행할 필요 없이 제공되는 결과 파일을 실행하여 일반적인 분석을 수행할 수 있다.

이 외에도 다양한 분석 패키지들이 있지만 표준화된 방법은 아직 없으며, 분석의 각 단계에서 어떤 프로그램, 어떤 세부 옵션을 선택하느냐에 따라 분석결과가 달라질 수 있다 (Hamp et al., 2009; Sun et al., 2012). 따라서 연구자들은 자신의 연구 여건에 맞는 분석경로와 옵션을 선택하고, 비교하고자 하는 모든 시료에 동일한 방법을 사용하는 것이 중요하다. Mothur 프로그램의 개발자인 Schloss는 분석 경험을 통하여 도출된 최상의 분석경로를 온라인 상에서 제공하고 있으며, 이는 수시로 갱신된다 (www.mothur.org/wiki/

Schloss_SOP). 이 Schloss SOP는 각 연구자가 자신만의 분석경로를 선택하는데 있어 유용한 기준이 될 수 있다.

3. 토양미생물 분야의 응용사례

파이로시퀀싱을 이용한 미생물다양성 및 군집구조분석을 토양에 응용해서 출판된 논문은 많지 않으나 증가하는 추세에 있다. 미농무성 농업연구청의 Acosta-Martinez et al. (2008)은 토양관리와 토지이용이 세균 다양성에 미치는 영향을 구명하고자 미 텍사스주의 해발 800-1,100 미터에 위치한 Hale county와 Floyd county의 평원에서 교란되지 않은 초지[초지 단작 (P)과 다양한 잔디 혼작 (CRP)]와 2개의 농경지 [목화-겨울보리-옥수수 윤작 (Ct-W-Cr)과 목화 연속 단작 (Ct-Ct)]에서 세균 다양성을 조사하였다. 3%의 비유사도 수준에서 3개의 다양성 추정치를 사용한 결과, OTU (대략 종 수준에 해당)는 4,500을 넘지 않았으며, 다양성은 목화-겨울보리-옥수수 윤작에서 가장 높았고 목화 단작에서 가장 낮았다 (Ct-W-Cr 〉 P = CRP 〉 Ct-Ct). 이 토양의 우점 문은 *Actinobacteria*, *Bacteroidetes*, *Fermicutes* 순이었다. *Bacteroidetes* 비율은 초지 시스템 (P와 CRP)에 비해 농경시스템 (Ct-W-Cr, Ct-Ct)에서 더 높았으나, *Actinobacteria*는 교란되지 않은 초지 단작 포장과 다양한 잔디 혼작 포장 (P와 CRP)에서 더 높았다. 연구팀은 파이로시퀀싱 분석이 세균의 분포뿐만 아니라 존재 유무의 측면에서, 토양 관리와 토지 이용에 따른 세균다양성의 특성 변화를 파악하는데 강력한 방법임을 역설하였다.

다른 연구에서는 반건조 지역의 가축 (소)-목화 생산 시스템이 토양 세균 개체군의 특성에 미치는 영향을 분석하고 목화연작과 비교하였다 (Acosta-Martínez et al., 2010b). 건조지대 작부체계와 경운이 토양 군집조성에 미치는 영향을 분석한 연구에서는 수수-겨울호밀 윤작 포장이 수수-목화 윤작, 목화-겨울호밀-수수 윤작 포장에 비해 *Bacteroidetes*와 *Proteobacteria*의 비율이 높고 *Actinobacteria*는 낮으며, *Chloroflexi*, *Gemmatimonadetes*, *Verrucomicrobia*는 무경운에 비해 경운 재배시 높았다 (Acosta-Martínez et al., 2010a). 결론적으로 미생물 군집은 경운의 이력보다는 작물 윤작에 의한 영향이 큰 것으로 보고하였다.

Eilers et al. (2010)은 저분자량 탄소기질이 세균 군집구조에 미치는 영향을 축소판시험 (microcosm)으로 수행하였다. 포도당, 글리신, 시트르산을 초지, 활엽수림, 침엽수림 토양에 첨가한 결과, 모든 탄소기질이 α-*Proteobacteria*, γ-*Proteobacteria*, *Actinobacteria*의 상대비율을 증가시켜, 전체 군집구조에 영향을 주었다. 특히 시트르산은 세균군집에 더 큰 영향을 주었다. 이러한 결과는 기질과 시험 토양에 따라 다를 수 있겠지만, 군집 수준에서 첨가 기질에 우선적으로 반응하는 특이적인 세균 분류군이 있다는 것을 의미한다.

대륙적 규모로 토양 세균 군집구조를 조사하는데 파이로시퀀싱이 이용되기도 하였다. Lauber et al. (2009)은 북아메리카, 남아메리카를 아우르는 88개 토양을 수집하여, 총 152,359개 (토양 당 평균 1,501개)의 염기서열을 이용하여 세균군집의 특성을 연구하였다. 전체 세균 군집 조성은 토양 pH와 고도로 유의성 있는 상관관계 (r=0.79)가 있으며, 이는 토양 pH 범위에 따라 *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*의 상대 풍부도의 변화에 의한 것으로 해석하였다. 계통분류학적 다양성은 중성 부근의 pH에서 최대값을 보였고, pH와 높은 상관관계 (R^2=0.50)가 있었다. 결론적으로 토양 세균 군집의 구조는 큰 공간 스케일에서도 어느 정도 예측이 가능하다고 하였다. 이 연구는 콜로라도 대학의 Noah Fierer 연구팀에 의해 수행되었으며, 2010년에는 pH 기울기가 있는 경작지를 대상으로 pH가 세균과 진균 군집에 미치는 영향을 보고하였다 (Rousk et al., 2010). 즉 pH가 4.0-8.3의 범위에 있는 영국 로담스테드시험장의 장기 석회시험구 (Hoosfield acid strip)에서 채취한 27개의 시료를 qPCR과 파이로시퀀싱으로 분석한 결과, 세균의 상대풍부도와 다양성은 pH와 정의 상관관계를 보인 반면 (다양성은 pH 8.0이 pH 4.0에 비해 2 배 높았음), 진균의 상대풍부도는 pH에 의한 영향이 작고, 다양성은 pH와 상관관계가 낮았다. 따라서 진균과 달리 세균 군집 구조가 형성되는데 pH가 큰 영향을 끼친다는 것을 재차 확인하였다.

유기농법와 관행농법의 미생물 군집 차이를 구명하기 위해 파이로시퀀싱을 이용한 연구 (Sugiyama et al., 2010)는 미국 콜로라도의 감자재배지에서 수행되었다. 진균의 18S ribosomal RNA 유전자를 분석한 결과, 유기농지는 관행농지에 비해 미생물군집의 다양성과 균등도가 다소 높고, OTU 추정치는 농법 간 차이가 관찰되지 않았다. 그러나 16개의 OTU 중 7개는 관행에 비해 유기농지에서 상대풍부도가 증가한 반면 9개는 감소하였다. *Alternaria*속의 상대풍부도는 관행농지에서 높았고 (30.2:7.8%), *Pythium ultimum*은 유기농지에서 더 높았다 (0.25:0.05%). *A. solani*, *Phoma foveata*, *P. ultimum*에 특이 프라이머에 의한 qPCR 결과는 파이로시퀀싱과 유사하였는데, 이는 파이로시퀀싱 분석으로 OTU의 상대풍부도를 정량할 수 있다는 것을 의미한다.

바이오숯 (biochar)을 토양에 시용하면 장기적으로 탄소가 격리되고 작물 생산성이 개선되기 때문에 친환경적이다. 바이오숯의 시용이 시설재배 파프리카의 근권 세균 군집구조에 미치는 영향에 관한 연구는 Kolton et al. (2011)에 의해 수행되었다. 총 20,142의 염기서열 중 92-95%는 *Proteobacteria*, *Bacteroidetes*, *Actinobacteria*, *Firmicutes*문으로 동정되었으며, 바이오숯을 처리함에 따라 *Bacteroidetes*문의 상대풍부도는 12%에서 30%로 증가한 반면, *Proteobacteria*는 71%에서 47%로 감소하였다. *Bacteroidetes*문 *Flavobacterium*속은 바이오숯에 가장 강하게 반응하여 상대 풍부도는 바이오숯을 처리 했을때 대조구 4.2%에서 19.6%로 증

가되었다. 그 외 바이오숯에 의해 키틴 분해균 (*Chitinophaga*), 섬유소 분해균 (*Cellvibrio*), 방향족화합물 분해균 (*Hydrogenophaga*, *Dechloromonas*)이 민감하게 반응하였다.

Cui et al. (2012)은 8년에서 500년까지 다양한 간척역사를 가진 농경지에서 파이로시퀀싱 분석을 통해 장기적인 세균 군집의 천이를 증명하였다. 해수 소택지가 농경지로 변하면서 세균군집구조는 초기 수십 년에 걸쳐 급격히 변화하였으며 그 이후는 완만하였다. 문 수준에서는 *Proteobacteria*의 감소와 *Verrucomicrobia*의 증가가 뚜렷하였고, 논의 경우는 *Chloroflexi*와 *Acidobacteria*의 증가가, 밭의 경우는 *Actinobacteria*와 *Firmicutes*의 증가가 관찰되었다. 또한 세균 다양성은 초기에 급격히 증가하는 것으로 나타났다. Ahn et al. (2012)은 57년간 동일 비료를 연용한 논토양의 세균 및 고세균 군집구조를 파이로시퀀싱을 이용하여 분석하였다. 그 결과, 논토양에서는 *Chloroflexi*, *Proteobacteria*, *Actinobacteria*, *Crenarchaeota*문이 우점하였고, 세균과 고세균의 군집구조는 계절과 시비형태에 주로 영향을 받는 것으로 나타났다. 특히 시비형태 자체보다는 그로 인한 토양 pH 변화가 세균과 고세균의 군집구조에 더 큰 영향 주었으며, 이 결과는 토양 pH가 원핵생물의 군집구조 변화에 중요한 인자라는 것을 다시 한번 입증하였다.

이상의 연구결과는 파이로시퀀싱 방법이 그 동안 널리 사용해왔던 유전자지문 분석 (DGGE, T-RFLP) 등에 비해 대량의 유전자 염기서열을 이용하기 때문에 분류 해상도가 더 높을 뿐 아니라 cloning-sequencing에 비해 시간과 비용이 절약되고 종 추정치, 다양성 지수, 균등도 지수 등의 신뢰성이 높은 데이터를 생성할 수 있다는 것을 증명하였다.

결 론

최근까지, 미생물 군집 분석은 이용 가능한 기술의 부족이 큰 장애요인이었다. 2005년 이후 차세대 염기서열 분석 기술의 출현으로 많은 연구자들은 다양한 서식지의 시료를 이용하여 폭발적으로 염기서열을 해독하고 있으며, 이 기술은 미생물 생태 분야에 가히 혁명적인 영향을 주고 있다. 염기서열 해독 비용은 감소되고 처리량은 증가되고 부가적인 기술이 진보함에 따라 미생물 생태 연구의 환경이 변화되고 있다. 그러나 아직도 우리는 토양미생물 다양성에 대해서 빙산의 일각을 보고 있을 뿐이며, 증폭산물의 오류, 차세대 염기서열 분석 기술 자체의 문제, 표준화된 분석경로의 부족 등의 해결해야 할 난제들이 남아 있다. 결론적으로 미생물 다양성연구에 사용되는 파이로시퀀싱을 통해 우리는 실험실 조건에서 배양될 수 없는 것을 포함하여 토양미생물에 대해 유례 없이 많은 정보들을 생산하고 있으며, 이 기술을 통해 미생물다양성에 대해 더 완전하고 편향되지 않은 시각을 가질 수 있을 것으로 전망한다.

사 사

본 연구는 농촌진흥청 공동연구사업 (과제번호: PJ006906052012)에 의해 수행되었음

인 용 문 헌

Acinas, S.G., R. Sarma-Rupavtarm, V. Klepac-Ceraj, and M.F. Polz. 2005. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Appl. Environ. Microbiol. 71:8966-8969.

Acosta-Martínez, V., S. Dowd, Y. Sun, and V. Allen. 2008. Tag-encoded pyrosequencing analysis of bacterial diversity in a single soil type as affected by management and land use. Soil Biol. Biochem. 40:2762-2770.

Acosta-Martínez, V., S.E. Dowd, C.W. Bell, R. Lascano, J.D. Booker, T.M. Zobeck, and D.R. Upchurch. 2010a. Microbial community composition as affected by dryland cropping systems and tillage in a semiarid sandy soil. Diversity. 2:910-931.

Acosta-Martínez, V., S.E. Dowd, Y. Sun, D. Wester, and V. Allen. 2010b. Pyrosequencing analysis for characterization of soil bacterial populations as affected by an integrated livestock-cotton production system. Appl. Soil Ecol. 45:13-25.

Ahn, J.H., J. Song, B.Y. Kim, M.S. Kim, J.H. Joa, and H.Y. Weon. 2012. Characterization of the bacterial and archaeal communities in rice field soils subjected to long-term fertilization practices. J. Microbiol. 50:754-765.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410.

Amann, R.I., W. Ludwig, and K.H. Schleifer. 1995. Phylogenetic identification and in-situ detection of individual microbial cells without cultivation. Microbiol. Rev. 59:143-169.

Ashelford, K.E., N.A. Chuzhanova, J.C. Fry, A.J. Jones, and A.J. Weightman. 2005. At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. Appl. Environ. Microbiol. 71:7724-7736.

Bartram, A.K., M.D.J. Lynch, J.C. Stearns, G. Moreno-Hagelsieb, and J.D. Neufeld. 2011. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. Appl. Environ. Microbiol. 77:3846-3852.

Benson, D.A., I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, and E.W. Sayers. 2011. GenBank. Nucleic Acids Res. 39:D32-D37.

Caporaso, J.G., J. Kuczynski, J. Stombaugh, K. Bittinger, F.D. Bushman, E.K. Costello, N. Fierer, A.G. Pena, J.K. Goodrich, J.I. Gordon, G.A. Huttley, S.T. Kelley, D.

Knights, J.E. Koenig, R.E. Ley, C.A. Lozupone, D. McDonald, B.D. Muegge, M. Pirrung, J. Reeder, J.R. Sevinsky, P.J. Tumbaugh, W.A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, and R. Knight. 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods. 7:335-336.

Caporaso, J.G., C.L. Lauber, W.A. Walters, D. Berg- Lyons, J. Huntley, N. Fierer, S.M. Owens, J. Betley, L. Fraser, M. Bauer, N. Gormley, J.A. Gilbert, G. Smith, and R. Knight. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 6:1621-1624.

Chenna, R., H. Sugawara, T. Koike, R. Lopez, T.J. Gibson, D.G. Higgins, and J.D. Thompson. 2003. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. 31:3497-3500.

Cole, J.R., Q. Wang, E. Cardenas, J. Fish, B. Chai, R.J. Farris, A.S. Kulam-Syed-Mohideen, D.M. McGarrell, T. Marsh, G.M. Garrity, and J.M. Tiedje. 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 37:D141-D145.

Cui, J., H. Meng, M. Nie, X. Chen, Z. Li, N. Bu, B. Li, J. Chen, Z. Quan, and C. Fang. 2012. Bacterial succession during 500 years of soil development under agricultural use. Eco. Res. 27:793-807.

DeSantis, T.Z., P. Hugenholtz, N. Larsen, M. Rojas, E.L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, and G.L. Andersen. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 72:5069-5072.

Dolfing, J., A. Vos, J. Bloem, P.A.I. Ehlert, N.B. Naumova, and P.J. Kuikman. 2004. Microbial diversity in archived soils. Science. 306:813-813.

Edgar, R. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 5:113.

Edgar, R.C., B.J. Haas, J.C. Clemente, C. Quince, and R. Knight. 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 27:2194-2200.

Eilers, K.G., C.L. Lauber, R. Knight, and N. Fierer. 2010. Shifts in bacterial community structure associated with inputs of low molecular weight carbon compounds to soil. Soil Biol. Biochem. 42:896-903.

Garrity, G.M., T.G. Lilburn, J.R. Cole, S.H. Harrison, J. Euzeby, and B.J. Tindall. 2007. The taxonomic outline of Bacteria and Archaea. TOBA Release 7.7. 10.1601 /TOBA7.7. Michigan State University Board of Trustees, Michigan, USA.

Gilles, A., E. Meglecz, N. Pech, S. Ferreira, T. Malausa, and J.F. Martin. 2011. Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. BMC Genomics. 12.

Haas, B.J., D. Gevers, A.M. Earl, M. Feldgarden, D.V. Ward, G. Giannoukos, D. Ciulla, D. Tabbaa, S.K. Highlander, E. Sodergren, B. Methé, T.Z. DeSantis, T.H.M. Consortium, J.F. Petrosino, R. Knight, and B.W. Birren. 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res. 21:494-504.

Hamp, T.J., W.J. Jones, and A.A. Fodor. 2009. Effects of experimental choices and analysis noise on surveys of the "rare biosphere". Appl. Environ. Microbiol. 75:3263- 3270.

Hongoh, Y., H. Yuzawa, M. Ohkuma, and T. Kudo. 2003. Evaluation of primers and PCR conditions for the analysis of 16S rRNA genes from a natural environment. FEMS Microbiol. Lett. 221:299-304.

Huber, T., G. Faulkner, and P. Hugenholtz. 2004. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics. 20:2317-2319.

Huse, S.M., J.A. Huber, H.G. Morrison, M.L. Sogin, and D.M. Welch. 2007. Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol. 8:R143.

Huse, S.M., L. Dethlefsen, J.A. Huber, D.M. Welch, D.A. Relman, and M.L. Sogin. 2008. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. PLoS Genetics. 4.

Huse, S.M., D.M. Welch, H.G. Morrison, and M.L. Sogin. 2010. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. Environ. Microbiol. 12:1889-1898.

Ishii, K. and M. Fukui. 2001. Optimization of annealing temperature to reduce bias caused by a primer mismatch in multitemplate PCR. Appl. Environ. Microbiol. 67:3753-3755.

Klammer, S., C. Mondini, and H. Insam. 2005. Microbial community fingerprints of composts stored under different conditions. Annals of Microbiology. 55:299-305.

Knight, R., J.G. Caporaso, C.L. Lauber, W.A. Walters, D. Berg-Lyons, C.A. Lozupone, P.J. Turnbaugh, and N. Fierer. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc. Natl. Acad. Sci. U.S.A. 108:4516-4522.

Kolton, M., Y.M. Harel, Z. Pasternak, E.R. Graber, Y. Elad, and E. Cytryn. 2011. Impact of biochar application to soil on the root-associated bacterial community structure of fully developed greenhouse pepper plants. Appl. Environ. Microbiol. 77:4924-4930.

Kunin, V., A. Engelbrektson, H. Ochman, and P. Hugenholtz. 2010. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. Environ. Microbiol. 12:118-123.

Lauber, C.L., M. Hamady, R. Knight, and N. Fierer. 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Appl. Environ. Microbiol. 75:5111-5120.

Lauber, C.L., N. Zhou, J.I. Gordon, R. Knight, and N. Fierer. 2010. Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated

samples. FEMS Microbiol. Lett. 307:80-86.

Li, W. and A. Godzik. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 22:1658-1659.

Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Förster, I. Brettske, S. Gerber, A.W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. König, T. Liss, R. Lüßmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K.H. Schleifer. 2004. ARB: a software environment for sequence data. Nucleic Acids Res. 32:1363-1371.

Peplies, J., R. Kottmann, W. Ludwig, and F.O. Glöckner. 2008. A standard operating procedure for phylogenetic inference (SOPPI) using (rRNA) marker genes. Syst. Appl. Microbiol. 31:251-257.

Polz, M.F. and C.M. Cavanaugh. 1998. Bias in template-to-product ratios in multitemplate PCR. Appl. Environ. Microbiol. 64:3724-3730.

Pruesse, E., J. Peplies, and F.O. Glöckner. 2012. SINA: accurate high throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics.

Qiu, X., L. Wu, H. Huang, P.E. McDonel, A.V. Palumbo, J.M. Tiedje, and J. Zhou. 2001. Evaluation of PCR- generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. Appl. Environ. Microbiol. 67:880-887.

Quince, C., A. Lanzen, R. Davenport, and P. Turnbaugh. 2011. Removing noise from pyrosequenced amplicons. BMC Bioinformatics. 12:38.

Ranjard, L., D.P.H. Lejon, C. Mougel, L. Schehrer, D. Merdinoglu, and R. Chaussod. 2003. Sampling strategy in molecular microbial ecology: influence of soil sample size on DNA fingerprinting analysis of fungal and bacterial communities. Environ. Microbiol. 5:1111-1120.

Rochelle, P.A., B.A. Cragg, J.C. Fry, R.J. Parkes, and A.J. Weightman. 1994. Effect of sample handling on estimation of bacterial diversity in marine sediments by 16S rRNA gene sequence analysis. FEMS Microbiol. Ecol. 15:215-225.

Rousk, J., E. Baath, P.C. Brookes, C.L. Lauber, C. Lozupone, J.G. Caporaso, R. Knight, and N. Fierer. 2010. Soil bacterial and fungal communities across a pH gradient in an arable soil. ISME J. 4:1340-1351.

Schloss, P.D. and J. Handelsman. 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl. Environ. Microbiol. 71:1501-1506.

Schloss, P.D. 2009. A high-throughput DNA sequence aligner for microbial ecology studies. PLoS ONE. 4:e8230.

Schloss, P.D., S.L. Westcott, T. Ryabin, J.R. Hall, M. Hartmann, E.B. Hollister, R.A. Lesniewski, B.B. Oakley, D.H. Parks, C.J. Robinson, J.W. Sahl, B. Stres, G.G. Thallinger, D.J. Van Horn, and C.F. Weber. 2009. Introducing mothur: open source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75:7537-7541.

Schloss, P.D. 2010. The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. PLoS Comput. Biol. 6:e1000844.

Schloss, P.D. and S.L. Westcott. 2011. Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. Appl. Environ. Microbiol. 77:3219-3226.

Scholz, M.B., C.-C. Lo, and P.S.G. Chain. 2012. Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. Curr. Opin. Biotech. 23:9-15.

Stackebrandt, E. and B.M. Goebel. 1994. A place for DNA-DNA reassociation and 16S ribosomal-RNA sequence-analysis in the present species definition in bacteriology. Int. J. Syst. Bacteriol. 44:846-849.

Sugiyama, A., J.M. Vivanco, S.S. Jayanty, and D.K. Manter. 2010. Pyrosequencing assessment of soil microbial communities in organic and conventional potato farms. Plant Dis. 94:1329-1335.

Sun, Y., Y. Cai, L. Liu, F. Yu, M.L. Farrell, W. McKendree, and W. Farmerie. 2009. ESPRIT: estimating species richness using large collections of 16S rRNA pyrosequences. Nucleic Acids Res. 37.

Sun, Y., Y. Cai, S.M. Huse, R. Knight, W.G. Farmerie, X. Wang, and V. Mai. 2012. A large-scale benchmark study of existing algorithms for taxonomy-independent microbial community analysis. Brief. Bioinform. 13:107-121.

Suzuki, M.T. and S.J. Giovannoni. 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. 62:625-630.

Torsvik, V., R. Sorheim, and J. Goksoyr. 1996. Total bacterial diversity in soil and sediment communities - A review. J. Ind. Microbiol. 17:170-178.

Torsvik, V. and L. Ovreas. 2002. Microbial diversity and function in soil: from genes to ecosystems. Curr. Opin. Microbiol. 5:240-245.

Tzeneva, V.A., J.F. Salles, N. Naumova, W.A. de Vos, P.J. Kuikman, J. Dolfing, and H. Smidt. 2009. Effect of soil sample preservation, compared to the effect of other environmental variables, on bacterial and eukaryotic diversity. Research in Microbiology. 160:89-98.

van der Heijden, M.G., R.D. Bardgett, and N.M. van Straalen. 2008. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. Ecol. Lett. 11:296-310.

von Wintzingerode, F., U.B. Gobel, and E. Stackebrandt.

1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS Microbiol. Rev. 21:213-229.

Wang, Q., G.M. Garrity, J.M. Tiedje, and J.R. Cole. 2007. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 73:5261-5267.

Watanabe, K., Y. Kodama, and S. Harayama. 2001. Design and evaluation of PCR primers to amplify bacterial 16S ribosomal DNA fragments used for community fingerprinting. J. Microbiol. Meth. 44:253-262.

Wright, E.S., L.S. Yilmaz, and D.R. Noguera. 2012. DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. Appl. Environ. Microbiol. 78:717-725.

Wu, G.D., J.D. Lewis, C. Hoffmann, Y.Y. Chen, R. Knight, K. Bittinger, J. Hwang, J. Chen, R. Berkowsky, L. Nessel, H.Z. Li, and F.D. Bushman. 2010a. Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. BMC Microbiol. 10:206.

Wu, J.Y., X.T. Jiang, Y.X. Jiang, S.Y. Lu, F. Zou, and H.W. Zhou. 2010b. Effects of polymerase, template dilution and cycle number on PCR based 16 S rRNA diversity analysis using the deep sequencing method. BMC Microbiol. 10:255.

Yoo, K., J. Lee, and J. Park. 2009. A review on the current methods for extracting DNA from soil and sediment environmental samples. J. Soil Groundwater Env. 14:57-67.

Zhou, H.W., D.F. Li, N.F.Y. Tam, X.T. Jiang, H. Zhang, H.F. Sheng, J. Qin, X. Liu, and F. Zou. 2011. BIPES, a cost-effective high-throughput method for assessing microbial diversity. ISME J. 5:741-749.