



Note: Soil/Pesticide/Environmental Sciences

Freeze-drying feces reduces illumina-derived artefacts on 16S rRNA-based microbial community analysis

Jungman Kim¹ · Tatsuya Unno¹

illumina를 이용한 16S rRNA 기반 미생물생태분석에서 분변의 동결건조에 의한 인공적인 시퀀스 생성 감소효과

김정만¹ · 운노타쯔야¹

Received: 29 June 2016 / Accepted: 25 July 2016 / Published Online: 31 December 2016
© The Korean Society for Applied Biological Chemistry 2016

Abstract When used for amplicon sequencing, Illumina platforms produce more than hundreds of sequence artefacts, which affects operational taxonomic units based analyses such as differential abundance and network analyses. Nevertheless it has become a major tool for fecal microbial community analysis. In addition, results from sequence-based fecal microbial community analysis vary depending on conditions of samples (i.e., freshness, time of storage and quantity). We investigated if freeze-drying samples could improve quality of sequence data. Our results showed reduced number of possible artefacts while maintaining overall microbial community structure. Therefore, freeze-drying feces prior to DNA extraction is recommended for Illumina-based microbial community analysis.

Keywords Artefacts · Fecal microbial community · Freeze-drying · Illumina

서론

차세대염기서열결정법(Next Generation Sequencing, NGS)은 오랫동안 분변미생물생태분석에 사용되고 있지만, NGS 기반 분변미생물생태분석에 대한 표준화된 방법이 없는 실정이다. Pyrosequencing은 Roche에서 이미 서비스가 중단 될 것이라고 알려졌음에도 불구하고 미생물생태분석을 위해 가장 많이 사용되고 있다. 하지만, Illumina사에서 제공하고 있는 MiSeq은 시간이 지남에 따라 가격 대비 효율적인 측면에서 우수해지고 있기 때문에 pyrosequencing을 대체하고 있다(Loman et al. 2012). Illumina 시퀀싱 플랫폼은 PCR 산물을 이용하여 시퀀싱이 진행되기 때문에 상당히 많은 양의 인위적인 시퀀스가 형성(Caporaso et al. 2011)이 될 뿐만 아니라 indel 및 substitution에 의해 오류를 야기시키는 문제점을 갖고 있으며(Dohm et al. 2008; Hoffmann et al. 2009; Kircher et al. 2009; Zhan et al. 2014), 또한 reverse reads가 forward reads보다 낮은 quality score를 갖고 있는 경향이 있다(Kwon et al. 2013). Schirmer (Schirmer et al. 2015) 연구팀은 MiSeq library 준비과정과 프라이머 선택은 오류가 일어날 수 있는 가장 중요한 요소라고 보고하였으며, 분변샘플을 이용한 MiSeq 준비과정에서 샘플상태 또한 2차적 오류 요소로 알려져 있다. 예를 들어, Liu (Li et al. 2007) 연구팀은 분변샘플의 균질화과정은 분변미생물생태 간 편차를 감소시킬 수 있는 요소라고 보고하였으며, 그와 더불어 DNA extraction method (Kennedy et al. 2014b), PCR template concentration (Kennedy et al. 2014a), PCR cycle condition (Ahn et al. 2012) 또한 분변미생물생태분석 결과에 영향을 미칠 수 있다고 보고 되었다. 이전 선행연구에서는 DNA 추출 과정에 bead-beating (Salonen et al. 2010)과 boiling samples (Peng et al. 2013) 과정이 포함되어 있으면 보다 효과적으로

Tatsuya Unno (✉)
E-mail: tatsu@jejunu.ac.kr

¹Faculty of Biotechnology, College of Applied Life Science, SARI, Jeju National University, Jeju 63243, Republic of Korea

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

DNA 추출이 가능하다고 제안하였다. 또한, 동결건조를 시킨 분변샘플로부터 추출한 DNA를 이용하여 *Escherichia coli* O157:H7 및 *Compylobacter jejuni*를 포함하는 병원성미생물의 검출 한계를 향상시킨다고 보고 되었다(Rapp et al. 2010). 이와 같이 분변샘플의 동결건조가 미생물 검출한계를 향상시켜준다는 연구결과가 보고 되었으나 분변미생물생태를 분석하는데 일반적으로 적용되고 있지 않다. 따라서, 본 연구에서는 분변샘플의 동결건조가 MiSeq 기반 미생물생태분석 결과에 어떠한 영향을 미치는지에 대하여 조사하였다.

재료 및 방법

샘플채취 및 DNA 추출

분변샘플은 제주에 위치한 농가에서 인위적으로 5개의 돼지분변샘플을 약 5 g씩 채취하였다. DNA는 약 250 mg의 신선한 분변(동결건조를 하지 않은 분변)으로부터 MOBIO Power Fecal DNA isolation kit (MOBIO Laboratories Inc., CA, USA)를 이용하여 추출하였으며, 남은 분변샘플들은 동결 시킨 후 약 2일 동안 동결건조과정을 거쳐 가루형태로 갈아 약 10 mg으로부터

동일한 DNA extraction kit를 이용하여 DNA를 추출하였다.

MiSeq library 제작

PCR은 이전 선행연구에서 제시한 방법으로 미생물/고세균의 16S rRNA 유전자의 V4 region을 증폭시켰다(Kozich et al. 2013). PCR은 Maxime PCR PreMix Kit (iNtRON Biotechnology Inc., Gyeonggi, Korea)를 이용하여 3반복 수행하였으며, 그 조건은 다음과 같다; 95 °C에서 2분간 initial denaturation 후, 95 °C에서 20초간 denaturation, 72 °C에서 15초간 annealing, 72 °C에서 1분간 extension을 30회 반복하고, 72 °C에서 5분간 final extension 하였다. PCR 산물은 AccuPrep Gel Purification kit (Bionner Inc., Deajeon, Korea)을 이용하여 정제한 후, Qubit assay (Invitrogen Inc., Carlsbad, CA, USA)를 이용하여 그 농도를 측정하였다. 모든 샘플의 PCR산물을 동일한 양이 들어가기도록 1개의 1.5 mL tube에 모아 시퀀싱을 보내기 전 까지 -20 °C에서 보관하였으며, Illumina MiSeq 플랫폼을 이용한 시퀀싱은 마크로젠(Macrogen Inc., Seoul, Korea)에서 수행하였다.

시퀀스데이터 프로세싱 및 분석

Paired-end 형식의 MiSeq 시퀀싱으로부터 각 샘플마다 2개의

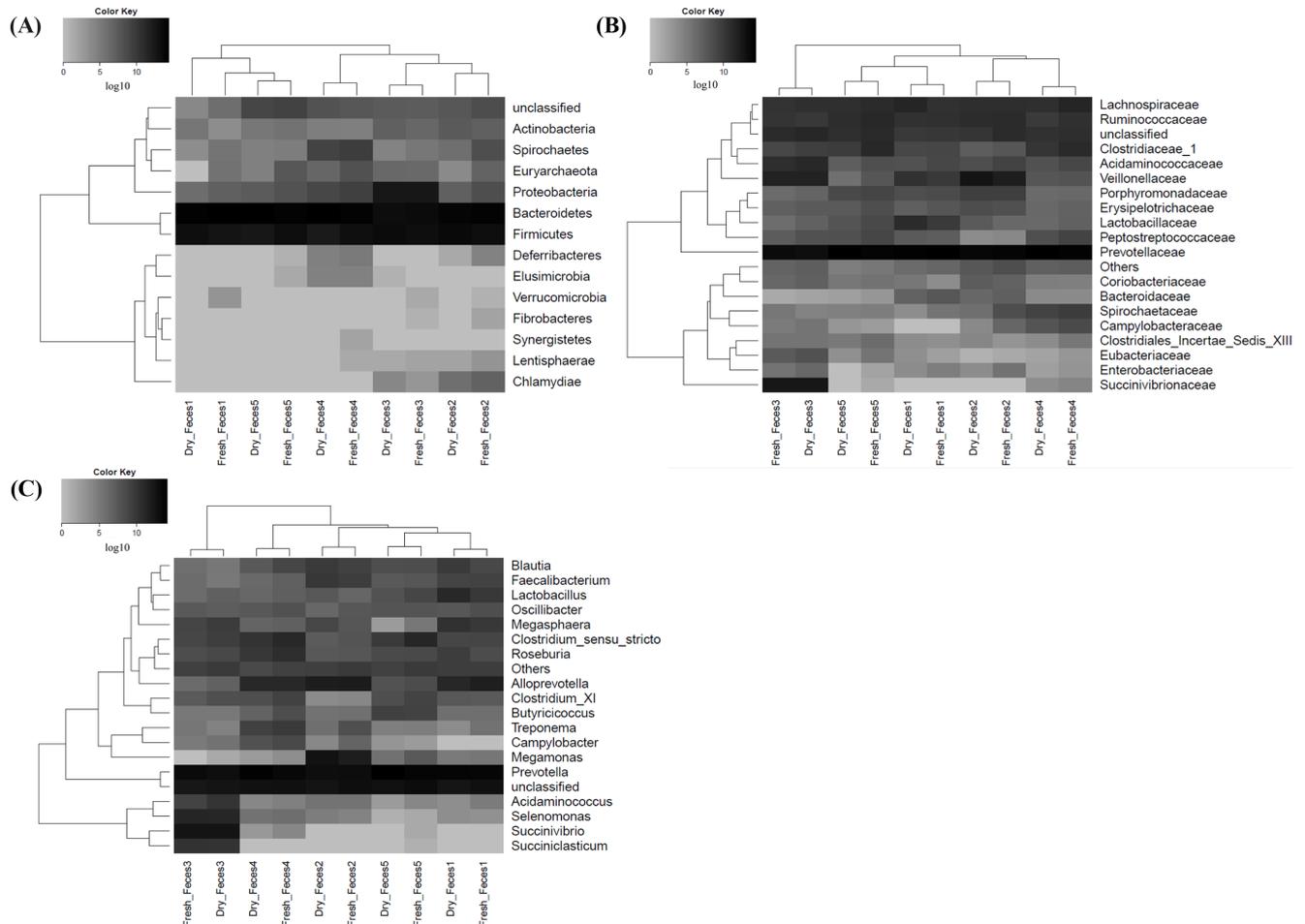


Fig. 1 Analysis of taxonomic composition at the phylum (A), family (B), and genus (C) levels.

Table 1 Portion of shared and unshared operational taxonomic units (OTUs) between microbial communities derived from freeze-dried and fresh feces

Sample	No. of read	Total OTUs	No. of shared OTUs	No. of unshared OTUs	No. of rare OTUs* in unshared OTUs	
Feces1	Fresh	36,203	256	50 (19.5 %)	38 (76.0 %)	
	Freeze dry	47,756	259	206	53 (20.5 %)	39 (73.6 %)
Feces2	Fresh	50,659	420	259	161 (38.3 %)	112 (69.6 %)
	Freeze dry	37,401	313	54 (17.3 %)	42 (77.8 %)	
Feces3	Fresh	36,714	332	226	106 (31.9 %)	84 (79.3 %)
	Freeze dry	33,672	301	75 (24.9 %)	55 (73.3 %)	
Feces4	Fresh	42,287	371	277	94 (16.6 %)	71 (75.5 %)
	Freeze dry	29,671	317	40 (12.6 %)	31 (77.5 %)	
Feces5	Fresh	37,191	340	258	82 (24.1 %)	71 (86.6 %)
	Freeze dry	37,471	301	43 (14.3 %)	33 (76.7 %)	

* Rare OTUs: OTUs having only singletons

fastq 파일을 얻었으며, PEAR 소프트웨어(Zhang et al. 2014)를 이용하여 assembly 과정을 거쳐 각각 하나의 fastq 파일로 합쳤다. Assembly된 fastq 파일은 Mothur 파이프라인(Schloss et al. 2009)에 따라 이후 분석을 수행하였다. 간단히 설명하자면 Assembled sequences는 SILVA rRNA database (Quast et al. 2013)를 이용하여 align과정을 거쳐, non-aligned sequences를 제거하였다. Chimeric sequences는 UCHIME (Edgar et al. 2011)을 이용하여 제거하였고, taxonomic classification는 Ribosomal Database Project (Cole et al. 2009) training set version 10을 이용하여 수행하였으며, 이를 통해 고세균 및 박테리아로 분류되지 않는 시퀀스를 제거하였다. 본 연구팀의 선

행연구에서 제안하였듯이 Singletons는 Mothur 서브루틴 “split.abund”를 이용하여 제거하였고(Unno 2015), Operational taxonomic units (OTUs)는 distance 0.03 (97 % sequence similarity)으로 계산하여 분류하였다. Mothur는 또한 richness와 evenness를 나타내주는 생태학적 지표를 얻기 위하여 사용하였으며, non-metric multidimensional scaling (NMDS)는 미생물 생태 간 차이를 비교하기 위하여 OTUs 기반으로 구축하였다. ANOVA (Analysis of variances) 및 AMOVA (Analysis of molecular variances)는 생태학적 지표 및 미생물생태를 비교하기 위한 유의성 검정방법으로 사용하였다.

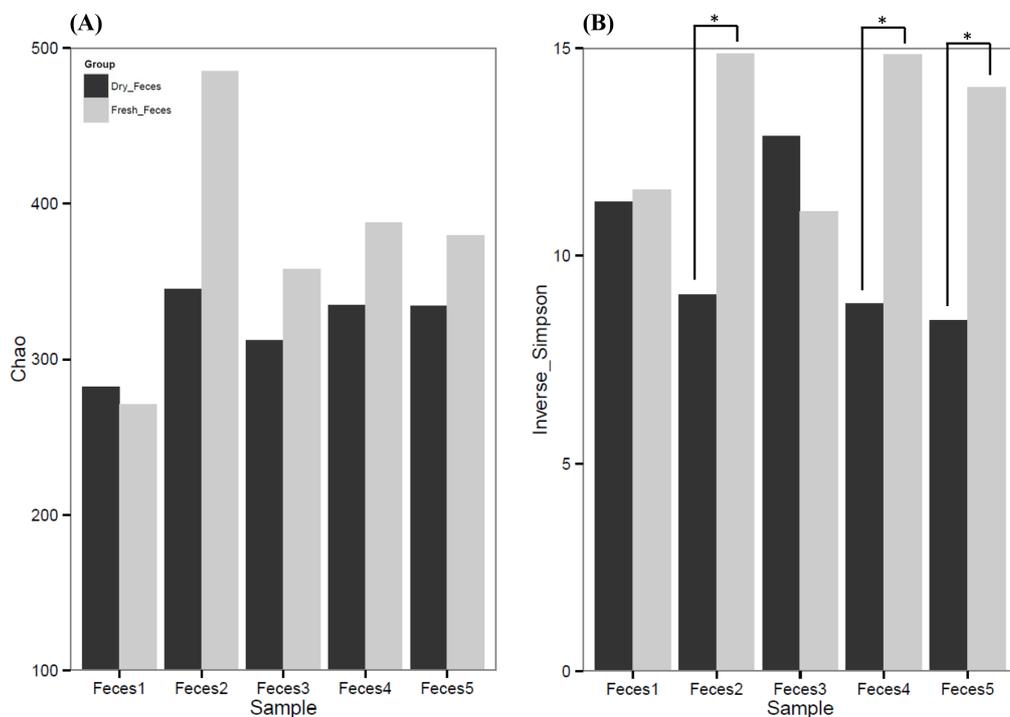


Fig. 2 Comparison of ecological indices in fecal microbiota derived from fresh or freeze-dried fecal materials: Species richness Chao (A) and species evenness inverse Simpson (B)

결 과

MiSeq 시퀀싱 데이터

Supplementary Fig. 1은 본 연구에서 사용한 모든 샘플로부터 충분한 수의 시퀀스 수를 얻었다는 것을 보여준다. 총 493,000 reads를 10개의 분변샘플(신선한 분변 F1-5, 동결건조한 분변 D1-5)로부터 얻었으며, 에러시퀀스 제거 프로세싱을 통해서 103,975 (21.1 %)가 제거되었다. 에러시퀀스와 singletons의 시퀀스 수가 Feces1과 Feces3를 제외하고 동결건조한 분변에서 보다 신선한 분변에서 더 높게 나타나는 것을 확인하였으며 (Supplementary Table 1), 이러한 결과는 신선한 분변에서 보다 많은 에러시퀀스가 형성된다는 것을 나타낸다. 최저 시퀀스 수가 29,671 reads로 확인되어 이를 이용하여 모든 샘플들의 시퀀스 수를 표준화시켰다. 본 연구에서 사용한 모든 시퀀스 데이터는 Short Read Archive (SRA)에 등록하여 Bioproject 번호 PRJNA326327을 부여 받았다.

동결건조 유무에 따른 분변간 OTUs 기반 미생물생태분석

신선한 분변과 동결건조한 분변 간 Taxonomic composition 차이를 확인하기 위해 phylum, family, genus 수준에서 조사하였다. Figure 1은 heat map analysis를 통해서 Relative abundance를 바탕으로 샘플간 유사성을 보여주는데 신선한 분변과 동결

건조한 분변 간 분변미생물생태가 유의적인 차이가 없는 것으로 확인되었으며, NMDS 결과인 Fig. S2에서 또한 두 그룹 간 미생물생태가 차이가 없음을 확인하였다. 본 결과는 동결건조과정의 전체적인 분변미생물생태에는 영향을 미치지 않는다는 것을 보여준다. 반면에, α -diversity 분석을 생태학적 지표로 나타낸 species richness (Chao)와 species evenness (Inverse Simpson)를 신선한 분변과 동결건조한 분변샘플간 비교분석을 실시하였다(Fig. 2). 신선한 분변과 동결건조한 샘플간 species richness는 유의적인 차이를 나타내지 않았지만(Fig. 2A), species evenness는 3개의 분변(D2, D4, D5)에서 유의적으로 낮게 나타나는 것을 확인하였다($p < 0.05$) (Fig. 2B).

동결건조 유무에 따른 나눈 동일한 분변샘플 간 unshared OTUs 비율을 Table 1에 정리하였는데, unshared OTUs가 동결건조한 분변샘플에서 보다 신선한 분변샘플에서 더 높게 나타나는 것을 확인하였다. 또한 low abundance (1~10)를 갖는 OTUs, 즉 rare OTUs 수를 read abundance 별로 분석을 실시하였다(Fig. 3). 그 결과, 신선한 분변에서 rare OTUs가 더 많이 존재하는 것으로 확인되었다. 따라서 분변샘플의 동결건조는 보다 적은 unshared OTUs 및 rare OTUs가 감소되며, 그에 따라 species evenness가 신선한 분변샘플에 비해 낮게 나타난 것이라고 할 수 있다.

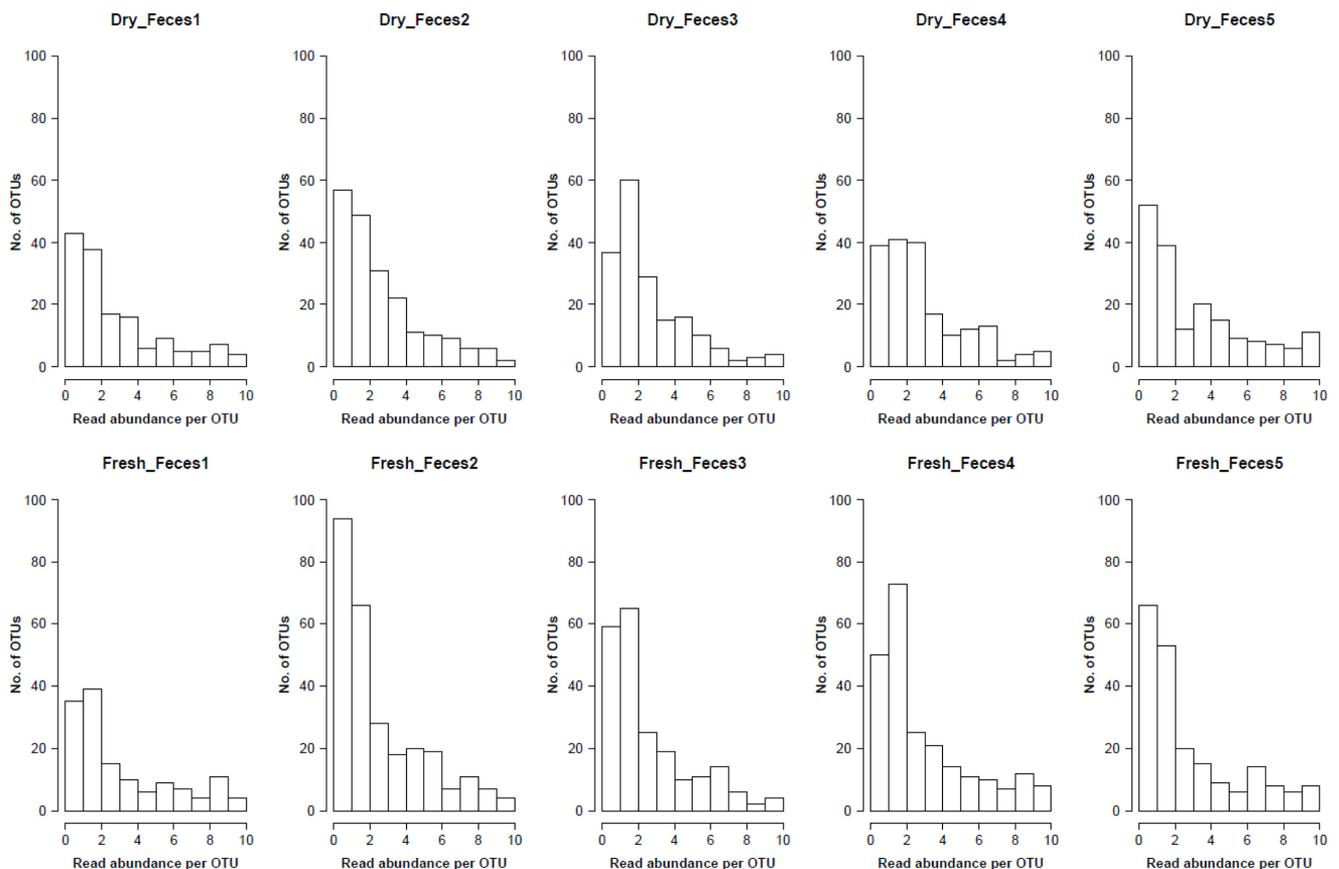


Fig. 3 Distribution of low abundance (from 1 to 10) operational taxonomic units in fecal microbiota derived from fresh or freeze dried feces

고찰

본 연구에서 신선한 분변 약 250 mg을 동결건조한 결과 dry weight가 약 10 mg이라는 것을 확인하여, DNA 추출을 위해 신선한 분변 250 mg과 동결건조한 분변 10 mg을 사용 하였다. MiSeq 기반의 미생물생태분석에 대한 수요가 점점 증가함에 따라 시퀀싱과정 중에 오류를 감소 시키는 것이 재현이 가능한 결과를 도출하는데 중요한 요소가 되었다. 많은 연구에서 생명정보학을 통해 오류적이고 낮은 퀄리티의 시퀀스를 확인하고 제거 할 수 있다고 보여주고 있지만, 연구실에서 다양하게 사용하고 있는 DNA extraction kit (Kennedy et al. 2014b) 및 PCR conditions (Ahn et al. 2012) 등은 미생물생태분석 결과에 영향을 미칠 수 있다고 보여주고 있다. 본 연구에서 신선한 분변과 동결건조한 분변 간 미생물생태조사를 통하여 전체적인 미생물생태는 변화하지 않았지만 unshared OTUs의 수가 신선한 분변에서 더 많이 나타나는 것을 확인하였으며, 이로 인해 species evenness가 상대적으로 증가하는 것으로 나타났다. Caporaso (Caporaso et al. 2011) 연구팀은 많은 수의 rare OTUs가 mock community 분석 결과에서 확인 되었다고 보고하였으며, higher sequence depth는 시퀀싱 과정 동안 생성 될 수 있는 인위적인 시퀀스를 보다 많이 생성할 가능성이 있다고 보고하였다(Kircher et al. 2009; Kwon et al. 2013). 그러므로 인위적으로 형성되는 시퀀스의 증가는 rare OTUs 증가를 통해 나타나며, 이는 species evenness에 직접적으로 영향을 준다는 것을 보여준다. DNA를 추출하기 이전에 분변 동결건조를 처리하는 것은 이와 같이 인위적으로 형성되는 시퀀스를 감소시킬 수 있을 것이라고 사료된다. 따라서, 본 연구에서는 분변샘플의 동결건조처리와 그 분변샘플을 가루형태로 만드는 과정은 MiSeq 기반 미생물생태분석에서 높은 퀄리티의 시퀀싱 결과를 얻을 수 있을 것이며, 이 뿐만 아니라 샘플의 장기보관, 균질화 과정, DNA 추출을 위해 일정한 샘플양을 사용하는 등 유용한 점을 제공하기 때문에 분변샘플을 동결건조 시켜 미생물생태분석에 사용하는 것이 효과적인 방법이라고 강력히 권고한다.

초록

PCR 산물을 이용한 시퀀싱방법 중 Illumina 플랫폼으로 시퀀싱을 수행하면 100개 이상의 인위적인 시퀀스가 생겨나며, 그러한 인위적으로 형성되는 시퀀스에 의해 Operational taxonomic units를 기반으로 한 미생물생태 변화 및 네트워크 분석에 영향을 미친다. 이러한 문제점이 있음에도 불구하고 분변미생물생태를 분석하는데 Illumina에서 제공하고 있는 시퀀싱을 주된 방법으로 사용하고 있으며, 또한 그러한 시퀀스 기반의 분변미생물생태분석 결과는 분변샘플상태(i.e., 분변 보관 기간, 분변양, 분변의 신선도)에 따라 상이하게 나타난다. 본 연구에서는 분변샘플의 동결건조가 시퀀스 데이터의 퀄리티를 향상시키는지 관해 조사하였으며, 이를 통해 분변샘플에 동결건조처리는 전체적인 미생물생태구조를 변화시키지는 않지만 인위적으로 형성되었을 가능성이 있는 시퀀스의 수를 감소시키는 것으로 확인되었다. 따라서, 분변으로부터 DNA를 추출하기 이전에 동결건조처리하는 방법을 Illumina 기반의 분변미생물생태분석에 사용하는 것

을 권장한다.

Keywords 동결건조 · 분변미생물생태 · 인공적으로 형성된 시퀀스 · 일루미나

감사의 글 본 연구는 농촌진흥청 연구사업(과제번호: PJ009782)의 지원에 의해 이루어진 것으로 연구비 지원에 감사 드립니다.

References

- Ahn JH, Kim BY, Song J, Weon HY (2012) Effects of PCR cycle number and DNA polymerase type on the 16S rRNA gene pyrosequencing analysis of bacterial communities *J Microbiol* 50:1071–1074 doi:10.1007/s12275-012-2642-z
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample *Proceedings of the National Academy of Sciences of the United States of America* 108 Suppl 1:4516–4522 doi:10.1073/pnas.1000080107
- Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis *Nucleic Acids Res* 37:D141–145 doi:10.1093/nar/gkn879
- Dohm JC, Lottaz C, Borodina T, Himmelbauer H (2008) Substantial biases in ultra-short read data sets from high-throughput DNA sequencing *Nucleic Acids Res* 36:e105 doi:10.1093/nar/gkn425
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection *Bioinformatics* 27:2194–2200 doi:10.1093/bioinformatics/btr381
- Hoffmann S, Otto C, Kurtz S, Sharma CM, Khaitovich P, Vogel J, Stadler PF, Hackermuller J (2009) Fast mapping of short sequences with mismatches, insertions and deletions using index structures *PLoS Comput Biol* 5:e1000502 doi:10.1371/journal.pcbi.1000502
- Kennedy K, Hall MW, Lynch MD, Moreno-Hagelsieb G, Neufeld JD (2014a) Evaluating bias of illumina-based bacterial 16S rRNA gene profiles *Appl Environ Microbiol* 80:5717–5722 doi:10.1128/AEM.01451–14
- Kennedy NA, Walker AW, Berry SH, Duncan SH, Farquarson FM, Louis P, Thomson JM, Satsangi J, Flint HJ, Parkhill J, Lees CW, Hold GL (2014b) The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing *PLoS One* 9:e88982 doi:10.1371/journal.pone.0088982
- Kircher M, Stenzel U, Kelso J (2009) Improved base calling for the Illumina Genome Analyzer using machine learning strategies *Genome Biol* 10:R83 doi:10.1186/gb-2009-10-8-r83
- Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform *Applied and environmental microbiology* 79:5112–5120 doi:10.1128/AEM.01043-13
- Kwon S, Park S, Lee B, Yoon S (2013) In-depth analysis of interrelation between quality scores and real errors in Illumina reads *Conf Proc IEEE Eng Med Biol Soc* 2013:635–638 doi:10.1109/EMBC.2013.6609580
- Li F, Hullar MA, Lampe JW (2007) Optimization of terminal restriction fragment polymorphism (TRFLP) analysis of human gut microbiota *J Microbiol Methods* 68:303–311 doi:10.1016/j.mimet.2006.09.006
- Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ (2012) Performance comparison of benchtop high-throughput sequencing platforms *Nature biotechnology* 30:434–439 doi:10.1038/nbt.2198
- Peng X, Yu KQ, Deng GH, Jiang YX, Wang Y, Zhang GX, Zhou HW (2013) Comparison of direct boiling method with commercial kits for extracting

- fecal microbiome DNA by Illumina sequencing of 16S rRNA tags *J Microbiol Methods* 95:455-462 doi:10.1016/j.mimet.2013.07.015
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools *Nucleic Acids Res* 41:D590-596 doi:10.1093/nar/gks1219
- Rapp D, Waller J, Brightwell G, Muirhead RW (2010) Lyophilization prior to direct DNA extraction from bovine feces improves the quantification of *Escherichia coli* O157:H7 and *Campylobacter jejuni* *Appl Environ Microbiol* 76:1686-1688 doi:10.1128/AEM.01866-09
- Salonen A, Nikkila J, Jalanka-Tuovinen J, Immonen O, Rajilic-Stojanovic M, Kekkonen RA, Palva A, de Vos WM (2010) Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis *J Microbiol Methods* 81:127–134 doi:10.1016/j.mimet.2010.02.007
- Schirmer M, Ijaz UZ, D'Amore R, Hall N, Sloan WT, Quince C (2015) Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform *Nucleic Acids Res* 43:e37 doi:10.1093/nar/gku1341
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities *Appl Environ Microbiol* 75:7537-7541 doi:10.1128/AEM.01541-09
- Unno T (2015) Bioinformatic Suggestions on MiSeq-Based Microbial Community Analysis *J Microbiol Biotechnol* 25:765–770
- Zhan A, Xiong W, He S, Macisaac HJ (2014) Influence of artifact removal on rare species recovery in natural complex communities using high-throughput sequencing *PLoS One* 9:e96928 doi:10.1371/journal.pone.0096928
- Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR *Bioinformatics* 30:614-620 doi:10.1093/bioinformatics/btt593