BMB
Reports

Invited Mini Review

Next-generation approaches to the microbial ecology of food fermentations

Nicholas A. Bokulich[1,2,3] *& David A. Mills*[1,2,3]*

[1]Department of Viticulture and Enology, [2]Department of Food Science and Technology, [3]Foods for Health Institute, University of California, Davis, CA 95616, USA

Food fermentations have enhanced human health since the dawn of time and remain a prevalent means of food processing and preservation. Due to their cultural and nutritional importance, many of these foods have been studied in detail using molecular tools, leading to enhancements in quality and safety. Furthermore, recent advances in high-throughput sequencing technology are revolutionizing the study of food microbial ecology, deepening insight into complex fermentation systems. This review provides insight into novel applications of select molecular techniques, particularly next-generation sequencing technology, for analysis of microbial communities in fermented foods. We present a guideline for integrated molecular analysis of food microbial ecology and a starting point for implementing next-generation analysis of food systems. [BMB Reports 2012; 45(7): 377-389]

INTRODUCTION

Fermented foods have constituted a significant proportion of the human diet globally since prehistory. These foods benefit consumers through enhanced nutritional content, digestibility, microbial stability, and in some cases detoxification (1-5), likely representing mankind's oldest means of food preservation. In addition, these foods often serve as vehicles for beneficial microorganisms that play a beneficial role in human health, as well as prebiotic substances, which promote the growth and health-modulating activities of beneficial microbes in the human body (6). These foods are woven into the cultural fabric of the people who create them, and microbial fermentation remains a prominent means of food processing, enrichment, and preservation to the current day. Given the continued importance of fermented foods to the global population, much effort has been put towards describing the microbial communities responsible for different food fermentations to better improve and manipulate their production and safety. Molecular methods, relying on detection of nucleic acid sequences, have replaced traditional, culture-based detection methods for microbial community profiling due to their greater speed and accuracy—as most culture media bias for the growth of specific organisms, leading to distorted perspectives of microbial communities, even in simple microbial systems, such as food fermentations (7, 8). Additionally, the stressful conditions of some food systems, particularly alcoholic fermentations, can induce a viable-but-not-culturable state in microorganisms, preventing culture-based detection (9, 10).

This review will discuss current advances in molecular technology for studying food fermentations worldwide and present strategies for continued studies of global fermentations. This will not be a comprehensive review of all molecular profiling techniques, but will contrast a few leading technologies and, In particular, focus on the cutting-edge DNA sequencing technologies that are revolutionizing the study of food microbial ecology (Table 1).

MOLECULAR TOOLS

Targeted tools

A targeted profiling method is any technique that detects a taxonomically defined group of microbes, e.g., all bacteria, a specific species, or a specific strain. A multitude of targeted techniques exist, but the two discussed below are the principal methods employed in modern microbial ecology, particularly of foods. The principal issue of these and all targeted methods is that they can only quantify populations targeted by the selected primer or probe, so cannot distinguish subpopulations (without using additional probes) and cannot detect non-target populations. Therefore, these are not comprehensive profiling techniques (i.e., they do not characterize a complete microbial ecosystem), but are used to track functionally important members of a given ecosystem.

Fluorescence *in situ* hybridization (FISH)

Fluorescence *in situ* hybridization (FISH) is a culture-independent, PCR-free technique for directly visualizing microbial cells in a sample. This method utilizes fluorescence-labeled oligonucleotide probes targeting specific DNA sequences, usually re-

*Corresponding author. Tel: +1-530-754-7821; Fax: +1-530-752-0382; E- mail: damills@ucdavis.edu
http://dx.doi.org/10.5483/BMBRep.2012.45.7.148

Received 21 May 2012

Keywords: Food fermentation, Microbial community profiling, Molecular profiling tools, Next-generation sequencing, TRFLP

Table 1. Comparison of molecular tools

Tool	Target size (bp)[a]	Sensitivity[b]	Advantages	Disadvantages
FISH	15-30	10^5-10^6	PCR-free, direct visualization *in situ*	Targeted technique
QPCR	100-150	10^1-10^3	Sensitive, quantitative	Targeted technique
DGGE	100-300	10^2-10^3	Single-nucleotide fragment separation	Technically difficult
TRFLP	1,000+	10^2	High-throughput, pseudoquantitative	Lower taxonomic resolution
Illumina				
GAIIx	35-150	10^8	Superior sequence coverage, lowest sequencing cost	Shorter fragment length, theoretically lower taxonomic resolution
HiSeq2000	35-150	10^9		
MiSeq	35-150	10^7		
454 Life Sciences				
FLX Titanium	300-400	10^6	Longer fragment length, theoretically higher taxonomic resolution	Low coverage, expensive
FLX+	500-600	10^6		

[a]Maximum target size in bp. [b]Sensitivity as limit of detection (cells ml^{-1}) (FISH, QPCR, DGGE, TRFLP) or reads per run (Illumina, 454). All values are current as of the time of writing.

gions of rDNA, unique for discrete microbial taxa. Multiple probes can be multiplexed to enable simultaneous detection and enumeration of several targets in a mixed community, provided each is labeled with a different fluorophore exhibiting a unique emission spectrum. Cells are fixed and permeabilized *in situ*, incubated in the presence of the probe(s) to facilitate hybridization, and then observed directly by fluorescence microscopy or counted via flow cytometry if suspended in a fluid. This technique has become popular for profiling microbial communities as it avoids the challenges and biases of culturing and PCR, and additionally allows observation of target cells within their native environment, a feature with exciting applications to food fermentations. The only major issues with using FISH for microbial profiling is that it is not a high-throughput method and is not an efficient means of enumerating cells compared to quantitative PCR (QPCR). This is because only a few probes may be used simultaneously to avoid overlapping emission wavelength, and counting technology is either low-throughput (microscopy) or high-cost (flow cytometry). For more information regarding the design and use of FISH probes, readers are directed to the review of Bottari *et al*. (11).

FISH has been used to monitor microbial communities in a number of fermented foods, including yeasts and bacteria in wine (12-14), beer (15), and cheese (16-18). Ercolini and coworkers (19) designed a FISH procedure for *in situ* observation of bacteria in Stilton cheese. By embedding the cheese in a plastic resin, microsections could be removed and hybridized, enabling visualization of the spatial distribution of several bacterial species throughout the cheese (19). To the authors' knowledge, this is the only use of this novel strategy for studying any fermented food, but FISH has been used for other studies of microbial distribution in a solid matrix, e.g., of oral biofilms (20). While FISH is not a stand-alone method for microbial profiling, it presents a compelling complementary method for visualizing biogeography of microbial communities directly in their native environment and deserves more attention for the study of fermented foods, particularly those employing heterogeneous, solid-state fermentation, e.g., cheese, vegetable, meat, or seafood fermentations.

Quantitative PCR (QPCR)

Quantitative PCR (QPCR, also known as real-time PCR) has become an extremely popular technique for food microbiology testing and microbial ecology studies. In this review we will focus on novel applications of QPCR for the study of food microbiology and how advances in bioinformatics are increasing the utility of this and other molecular techniques. For a more thorough review of general methodology and applications of QPCR to food microbiology and food testing, the reader should consult the recent review of Postollec *et al*. (21).

QPCR incorporates the same principles of DNA replication used by PCR, but amplification is tracked in real-time using a fluorescent reporter molecule. A baseline threshold is determined, at which sample fluorescence can be distinguished from background noise. This is used to determine the threshold cycle (C_t) for each sample, at which fluorescence crosses the baseline threshold. A standard curve is constructed plotting C_t against gene copy number or cell count, and used to quantify unknown samples based on C_t. The result is an elegantly simple, rapid, and accurate means of tracking target cell or gene counts in samples.

QPCR has seen widespread use for quantification of microbial populations in foods (21) and can be used to profile complex, mixed-culture fermentations. Beyond cell quantification, the detection of functional genes presents a novel use of this technique with interesting applications to food microbiology. For example, this technique has been used to detect biogenic amine-producing strains of lactic acid bacteria (LAB) in wines (22, 23), ciders (24), and cheeses (25, 26) by directly targeting the genes responsible for biogenic amine production. Ibarburu and coworkers (27) have developed an assay to quantify exopolysaccharide (EPS)-producing LAB (which cause slime-production, a defect in wines and other beverages) by targeting the glycosyltransferase gene controlling EPS synthesis. QPCR of sulfite-reductase genes in *Saccharomyces cerevisiae* (28) can be used to track sulfide

production, a major off-flavor concern in wine fermentations. QPCR has also been used to detect beer-spoilage LAB by targeting the *horA* gene, which promotes resistance to hop antimicrobial iso-α-acids (29). This general approach detects populations based not on taxonomic distinction, but instead by specific genes with functions relevant to food properties. Another novel application of QPCR is the use of strain-specific primers/probes (30). Provided a strain-specific marker can be identified, this enables high-throughput, sensitive profiling of strain-level successions in food fermentations (30), which cannot be characterized by most culture-independent molecular profiling tools. Strain-specific QPCR is likely to increase as the advent of next generation sequencing (NGS) technologies has enabled researchers worldwide to rapidly generate draft genome sequences for any target microbial strain.

Community profiling tools

Culture-independent, molecular techniques for microbial profiling employ universal primers targeting a taxonomically defined group (usually kingdom-level phyla) to amplify heterogeneous sequences of DNA from mixed microbial samples before separating and taxonomically classifying the sequences, giving a perspective of the complete community structure of the target population. The common issue with these techniques is that they can only provide a qualitative or pseudoquantitative (i.e., relative abundance) assessment of a microbial community, so must be paired with quantitative, targeted molecular techniques to enumerate cells (i.e., absolute abundance). While a large number of profiling methods exist, this review will focus on the three most popular methods for microbial community profiling used today.

Denaturing gradient gel electrophoresis (DGGE)

Denaturing gradient gel electrophoresis (DGGE) has received a great deal of attention from food microbiologists in the past 20 years since it was first adopted for studying bacterial communities (31). DGGE and its close relative, temperature gradient gel electrophoresis (TGGE), function by separating short fragments of DNA along a chemical or temperature gradient, respectively. DNA is propelled through a polyacrylamide gel matrix by an electric current until it reaches the point along the gradient at which it becomes destabilized and partially denatures (held together by a high-stability GC-motif, or "GC tail"), becoming immobilized in the gel. The most abundant of these fragments may then be visualized on the gel under UV light, physically extracted, reamplified, and sequenced to qualitatively determine the most dominant organisms in a given sample.

In spite of its widespread popularity, DGGE is hindered by several issues that make it unsuitable for large-scale studies of microbial ecology. First of all, DGGE is a time-consuming and technically challenging method, as the denaturing gradient gels are difficult to set and run properly. DGGE only serves to separate DNA fragments, and a representative band from each position on a gel must be extracted and sequenced to identify the closest match. However, only a small number of samples may be run on a single gel, and so representative bands must be sequenced from each, making analysis of large sample sets time-consuming, redundant, and expensive. Band co-migration (32, 33) and rRNA gene multi-copy heterogeneity (34) further complicate this process, respectively hindering band purity and increasing redundancy. In addition, as a PCR-based method, DGGE is subject to PCR-induced error, including amplification bias (35) and chimera formation (36), distorting community structure. Another problem with this method is that it is non-quantitative, so can detect the presence of the dominant species present but cannot reliably determine relative abundance, though others dispute this point, suggesting integration of band size and intensity against a densitometric curve (37) may be an accurate indicator of abundance.

That said, DGGE was the tool that pioneered the field of molecular ecology of foods. DGGE was first implemented to study bacterial communities in food by Ampe *et al.* (8), who used it to describe the special distribution of bacteria in pozol, Mexican fermented maize dough. Cocolin and coworkers (38) introduced DGGE characterization of fungi, profiling yeast succession during wine fermentations. Prokaryotic 16S rRNA genes may contain insufficient heterogeneity to reliably identify some bacterial species using the short amplicons necessary for DGGE separation, and Renouf and coworkers (39) applied DGGE of bacterial *rpoB* gene amplicons as a means of differentiating wine LAB with high 16S rRNA gene homology. DGGE has seen widespread use in wine and food microbial community studies beyond enumeration here. For more information on the application of DGGE to study food, readers are referred to the review of Ercolini (40).

Terminal restriction fragment length polymorphism (TRFLP)

Terminal restriction fragment length polymorphism (TRFLP) is another extremely popular technique for studying microbial ecology, but has received much less attention by food microbiologists. This method differentiates microbial populations based on terminal restriction fragment (TRF) size. Mixed DNA samples are amplified with fluorescence-labeled universal primers, digested with select restriction enzymes, and separated by capillary electrophoresis with molecular standards for estimating fragment size. Only the labeled 5′ terminal fragments are detected and compared to a database to determine which bacterial populations (grouped as operational taxonomic units, OTUs) are represented by these fingerprint-like markers. Fluorescence intensity is integrated to pseudoquantitatively determine OTU abundance relative to total fluorescence.

Like DGGE, TRFLP is prone to the same amplification-related biases common to all PCR-based techniques. In addition, as electrophoretic mobility is a function of molecular weight, charge, and sequence structure, both sequence variability and fluorescent labels may alter the electrophoretic migration of TRFs, resulting in a reproducible disparity between expected and observed TRF sizes (41, 42). This problem may be ameliorated by

creating an empirical TRFLP database for the microbial species expected in a sample, but this process is time-consuming and can make critical assumptions about the microbial constituents of a given system. Residual polymerase activity (43) and single-stranded "pseudo-TRFs" (44) can also create artificial OTUs, though both of these errors can be corrected by proper procedural modification. The largest issue with TRFLP, however, is that sequence homology among related microbial taxa limits the taxonomic depth achieved using this technique, so a single TRF is identified to the common taxon encompassing all possible database matches. The shallow taxonomic depth often achieved by universal primers may be increased to species-level by using clade-specific primers, for example specific for LAB (45).

TRFLP has several advantages that suit it for microbial community profiling compared to DGGE. Most importantly, TRFLP is a high-throughput, automatable process, unlike DGGE, so can compare much larger sample sets rapidly and efficiently. Thus, TRF profiles may be integrated and used for diversity comparisons between samples on a massive scale. A common approach is to use principle component analysis (PCA) or a similar approach to cluster samples based on TRF abundance and visualize these differences in 2-dimensional space (46). Similarly, TRFLP is commonly used to calculate diversity statistics such as Jaccard distance (47) (though caution is advised, as TRFLP can overestimate species diversity due to the above-mentioned errors (48)) and ANOVA or other multivariate statistics can compare significance among clusters or treatment categories (46). Additionally, TRFLP is a pseudoquantitative method that calculates relative species abundance based on sensitive capillary electropherosis separation and fluorescence detection, whereas DGGE is not, based on polyacrylamide gel separation and UV detection. For interested readers, others have written excellent reviews detailing TRFLP methodology (42) and analysis (49) beyond the scope of this article.

TRFLP was first developed to study mixed bacterial communities by Liu and coworkers (50). Since then, it has become a popular technique for microbial profiling (49), but has been little-used for studying food systems. Rademaker and coworkers (51, 52) first applied TRFLP to study cheese and yogurt fermentation communities. Others have applied reverse-transcription TRFLP to analyze viable bacterial communities during cheese fermentations (53). Specific TRFLP assays have been developed for species-level differentiation of LAB (45) and yeasts (54) in fermented foods and beverages, and used to profile microbial successions during beer (55) and wine fermentations (56). TRFLP has also been used to analyze the bacterial composition of commercial probiotic products (57). This technique shows much promise as a low-tech, high-throughput method for studying microbial ecology, but has been largely overlooked by food microbiologists. With the advent of next-generation sequencing tools, TRFLP remains a viable complementary method for exploring species-level dynamics.

Next-generation sequencing (NGS)

Next-generation sequencing (NGS) collectively describes several technologies that achieve massively parallel sequencing of heterogeneous DNA fragments. For the purposes of microbial community surveillance, these fragments consist of short segments amplified using universal PCR primers targeting known marker genes, principally prokaryotic 16S rRNA and fungal ITS genes. Fragments are physically partitioned and base incorporation in the growing DNA strand is detected using novel technologies unique to each system. The only two NGS systems currently used for microbial community profiling (to our knowledge) are 454 Life Sciences pyrosequencing (58) and Illumina (formerly Solexa) sequencing platforms (59). These systems deliver a functionally identical product, raw DNA sequence, but with varying error rates, sequence lengths, yields, and costs. Overall, pyrosequencing can sequence longer reads (up to 600 bp) than Illumina systems (150 bp), but Illumina provides magnitudes-greater sequence coverage per run ($\sim 10^9$ reads on the HiSeq2000) than pyrosequencing ($\sim 10^6$ reads on the 454 Life Sciences Titanium FLX+), enabling greater multiplexing capacity and thus dramatically lower cost per sample (60). Longer sequencing reads theoretically provide greater taxonomic information, but reads as short as 100 bp can adequately describe microbial community changes and sequence coverage is more important for fully capturing microbial diversity (61).

Pyrosequencing, as the first commercially available NGS technique, has seen the most extensive use in food microbiology. It has been used to study the bacterial diversity of a number of fermented foods, including narezushi (fermented fish and rice) (62, 63), nukadoko (fermented rice bran) (64), Chinese liquor fermentations (65), Danish raw milk cheeses (66), Polish Oscypek cheese (66), and pearl millet fermentations (67). It has been remarkably well-utilized for studying Korean fermented foods, including bacteria and archaea of fermented seafood (68), bacterial and fungal communities of Makgeolli (rice beer) (69), meju (70) and doenjang (soybean pastes) (71), kochujang (fermented red pepper condiment) (72), and kimchi (73). It has also been used to study the viral communities inhabiting a range of Korean fermented food products (74). The greater sequence coverage and lower cost of Illumina sequencing has prompted its growing application in the study of microbial ecology in favor of pyrosequencing, and food microbiology is no exception. We have compared multiple hypervariable 16S rDNA domains for optimal taxonomic coverage of wine and food fermentations using short, 150-bp reads currently available on Illumina sequencing platforms, determining V4 as the more informative region (56). We have used this method to study botrytized wine fermentations (56) and the brewhouse-resident microbiota responsible for the fermentation of American coolship ale, a beer fermented without any inoculation (55).

The benefits of NGS over first-generation profiling methods, such as DGGE and TRFLP, are numerous. All NGS systems sequence thousands to billions of raw reads in a single run, translating to numerous observations for robust, saturated diversity

analyses. As detection occurs at the molecular level using high-resolution optics or a semiconductor chip, as opposed to the bulk UV/fluorescence excitation of DGGE and TRFLP, it is much more sensitive to low-abundance OTUs. Like TRFLP, the readout is pseudoquantitative (as read count per OTU relative to total reads), but it is sequence-based, so can be used for phylogeny-based diversity comparisons, such as UniFrac (75). Count-based diversity metrics, such as Bray-Curtis dissimilarity, calculate sample similarity as a function of shared OTUs, but do not consider phylogenetic distance, which often governs similarity in metabolic function and ecological role (75). Therefore, NGS enables comparison of communities by phylogenetic similarity and by extension functionality, in addition to count-based metrics. In food systems, this can give deeper meaning to observed microbial successions, demonstrating the impact of conditions (e.g., temperature, pH) or treatments not only on microbial diversity counts, but also on functional processes.

NGS is a growing technology and along with its many advantages come new challenges for analysis. The foremost issue is computational analysis, power, and storage. The many thousand to billion reads retrieved by NGS platforms come at the cost of many gigabytes-worth of hard drive space and enough computational power to handle these files. These requirements grow in parallel with increasing read coverage, so that the Illumina HiSeq2000 (~10 billion reads) requires magnitudes-greater space and memory than the 454 Life Sciences Titanium FLX pyrosequencer (~100,000 reads). Several open-source programs exist for processing short-amplicon NGS data, such as QIIME (76) and Mothur (77). These programs each package several features for the many steps required to translate raw reads into microbial community characterization, including high-throughput read filtration/demultiplexing, OTU selection, taxonomic assignment, sequence alignment, and downstream analysis. For a thorough discussion of the various programs available, their features, and the bioinformatic processes involved in interpreting microbial community patterns from NGS data, the reader is referred to the reviews of Kuczynski *et al*. (60) and Zaneveld *et al*. (78).

Database quality

A critical step for translating NGS reads into meaningful biological results is taxonomic assignment of sequences against a reference database. The resulting information is only as good as its reference database, so researchers must carefully select a database that meets the needs of their project. A database should be both comprehensive and high-quality. A comprehensive database should contain enough representative sequences to reliably describe all target taxa in thorough detail. If a given species is not represented within the database, sequences derived from that species would receive an incorrect assignment or remain unclassified. Therefore, a good database should have reasonable coverage of species of all taxa of interest.

Database quality is judged by the individual sequences contained therein. Many public sequence databases contain incorrectly classified or unclassified sequences, chimera, and sequences that are either too short or contain too many ambiguous base calls to be considered taxonomically useful (79-81). Such sequences can misconstrue NGS sequence assignment and/or limit depth of taxonomic classification, so users must carefully select databases that have been properly curated and filtered to include only high-quality, validated sequences. Fortunately, a number of options exist. For Prokaryotes, the Greengenes database (81, 82), Silva reference database (83), and Ribosomal database project (84) all provide curated, high-quality 16S rDNA sequence databases. For fungi, the situation is much less straightforward, as public sequence databases tend to contain very high levels of poor-quality and misannotated sequences (79, 80), while curated databases lack coverage or cover only specific subsections of fungi. UNITE (80, 85) offers a hand-curated ITS database of primarily ectomycorrhizal fungi (not associated with food) as well as a non-curated compilation of all ITS sequences from the International Nucleotide Sequence Database (INSD). Silva reference database (83) also contains a number of fungal large-subunit (LSU) rDNA sequences, but the list of curated sequences is far from complete. Different reference databases yield different taxonomic assignments as a function of completeness and quality (81, 86) so care must be taken to select an appropriate database for NGS studies, underlining the need for a comprehensive, carefully annotated fungal reference database (80).

Combined tactics

No molecular method is a stand-alone tool for describing microbial communities. Targeted methods can only detect and enumerate taxonomically defined targets, so cannot provide a comprehensive view of a community. Profiling methods, conversely, differentiate their target taxa, but cannot absolutely quantify these populations. Different molecular targets can also render dramatically different views of the same community due to primer bias (35), primer coverage (see general considerations below), and the degree of intra-taxon homology within that region (87). Even the cutting-edge profiling technology, NGS, cannot reliably differentiate most species, so cannot be considered fully comprehensive if species-level dynamics play a significant role in the studied system (as is usually the case for food fermentations).

Therefore, a combined approach is the best current strategy to study microbial ecology, utilizing multiple molecular tools to infer microbial patterns and pair these with quantifiable functional attributes impacting food fermentations. The combination of tools used will depend upon the questions asked and the system studied, but Fig. 1 presents a conceptual guideline for a combined approach to fermentation profiling. The numbered paragraphs below correspond to the steps outlined in Fig. 1.

1. Initially, a profiling tool should be used to give a broad-spectrum view of the microbial communities present. NGS would provide the most comprehensive, informative readout, but universal TRFLP or DGGE may be more appropriate given the system, sample size, and purpose. Microbial profiles may be used to calculate diversity statistics within and between samples

and cluster samples based on similarity. Optionally, a low-cost profiling tool such as TRFLP or DGGE may be used to cluster samples based on profile similarity. Then representative samples may be selected from each cluster for NGS, giving a more comprehensive, phylogenetically robust view of the microbial community (55).

2. A clade-specific profiling tool may be used to increase taxonomic depth for the most dominant and important groups. For example, LAB-TRFLP (45) may be used to profile LAB dynamics at species-level. Alternatively, a separate target region may be analyzed by the same profiling technique in step 1, confirming and/or complementing initial observations.

3. Profiling data should be complemented by QPCR or an equivalent quantitative, targeted method to quantify total bacterial/fungal load and/or track the most abundant or significant taxa detected by profiling tools.

4. If *in situ* visualization of microbial dispersion and/or physical interactions is an important aspect (e.g., in solid-phase fermentations), FISH may be used to observe spatial distribution of primary populations identified by profiling methods and QPCR. Representative samples may be selected by clustering based on community profiles (see item 1).

By pairing profiling and targeted tools to a specific purpose, a complete, reliable picture of whole-community dynamics may be formed, providing deeper insight into community function and process variables. In addition, further downstream steps (beyond the scope of this review) may be integrated with microbial data to form a system-wide perspective. The possibilities are virtually endless, but some examples given in Fig. 1 are:

5. Complement community-wide profiling with a strain profiling technique to observe strain-level dynamics during fermentation. This may in turn inform further targeted profiling (QPCR, FISH) using strain-specific primers/probes.

6. Complement microbial profiling with relevant physiochemical analyses (e.g., primary substrates, metabolites, volatile compounds).

7. Quantify functional gene presence/expression with QPCR to classify/compare communities by specific metabolic function. Alternatively, representative samples selected based on microbial profiles (as described in step 1) may be profiled using metatranscriptomic NGS to compare comprehensive expression profiles between different communties. Metatranscriptomics is another novel application of NGS technology beyond the scope of this discussion, and interested readers are referred to the review of Simon and Daniel (88).

8. Statistically integrate microbial, physiochemical, and func-

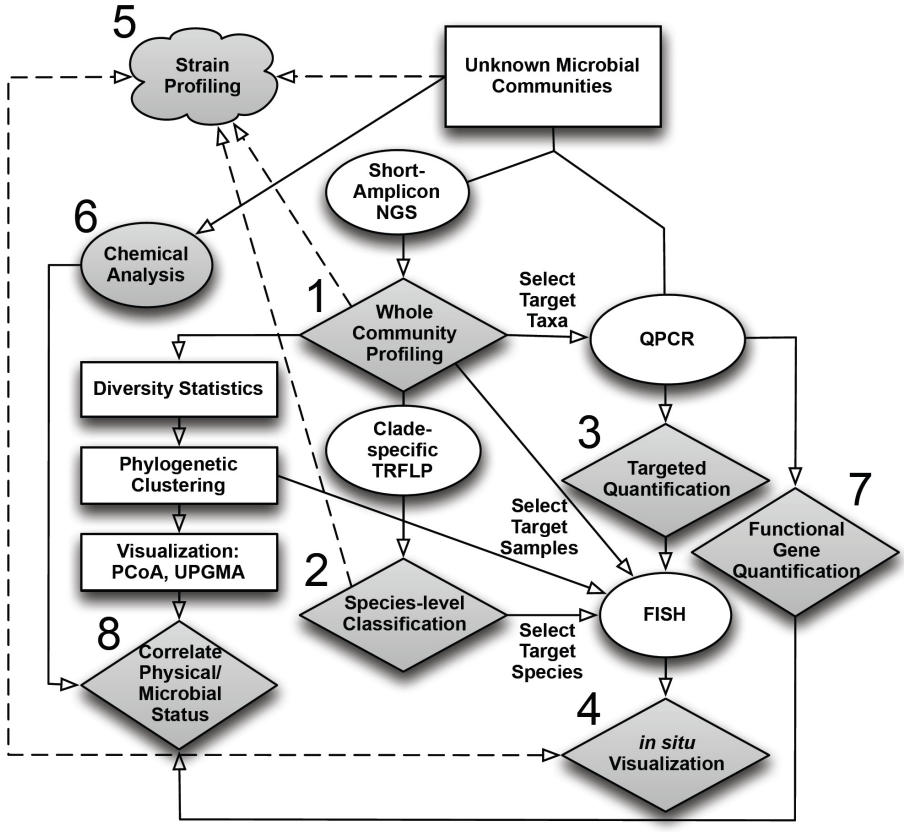

Fig. 1. General guideline for combined molecular approach to community profiling. General steps include: 1) broad-spectrum community profiling; 2) clade-specific profiling; 3) quantification of major taxa; 4) *in situ* visualization; 5) strain profiling; 6) chemical profiling; 7) functional gene quantification; and 8) integrate physiochemical and microbiological profiles to statistically correlate systems-level phenomena.

tional genetic profiles to determine how sample categories (e.g., treatment variables) impact microbiota and hence composition, and/or correlate physiochemical/functional genetic parameters with specific groups/strains of microbes.

GENERAL CONSIDERATIONS FOR MOLECULAR TOOLS

Sampling

Fermented food systems are almost always heterogeneous mixtures, often combining multiple solid, liquid, and gas phases. Even in beverage fermentations, tanks are usually not mixed, and microbial and abiotic components can often stratify or associate with phase interfaces (e.g., microbial films on solid surfaces, pellicle formation at points of air contact). If testing a commercial product *in situ*, these strata cannot be mixed, as this may disrupt the quality of the product itself. Careful strategies must be employed to ensure that a reasonably representative sample is taken while respecting the needs of the producer. Moreover, sample homogenization may become an issue for reliable sampling of solid and mixed-phase fermentations, wherein bias may be introduced by inadequate homogenization. Subsequent storage may also alter microbial communities, so appropriate handling and prompt processing are critical (89).

DNA extraction

Sample preparation and DNA extraction present further sources of sampling bias. Inadequate DNA sampling methods can alter community representations via incomplete lysis of particularly durable cells, and brief cell homogenization and/or enzymatic degradation steps can enhance extraction efficiency (90). Additionally, a number of food components, including lipids and proteins, may inhibit PCR amplification, and must be removed during extraction (91).

Viable cells

Food fermentations represent active, changing environments, and accurate assessment of the most dominant populations at a single time may require specific detection of viable cells. Reverse-transcription methods have been used frequently for viable cell detection with a large number of profiling methods (e.g., TRFLP) (53). As RNA exhibits lower stability than DNA, this technique can successfully capture viable cell activity through reverse-transcription of RNA transcripts, using cDNA instead as an amplification template for any PCR-based profiling method. However, RNA expression depends upon the physiological status of the cell, (92) and reverse-transcription techniques may not be sufficient for profiling communities of metabolically active cells during stationary phase. Another approach is to use the DNA-binding dyes propidium monoazide (PMA) or ethidium monoazide (EMA) prior to DNA extraction. These compounds covalently bind DNA, preventing PCR amplification, but cannot permeate the membrane of intact cells, so selectively bind DNA in nonviable cells (92). PMA has been successfully used to profile bacterial populations in environmental water samples using pyrosequencing (93) and yeast communities of wine fermentations using QPCR (94).

Molecular targets

Judicious DNA target selection is critical to the success of any molecular profiling technique. An optimal region will be chosen based on the limitations of the technique used, the degree of taxonomic discrimination required, and the system studied. Furthermore, primers or probes targeting this region must possess a reasonable balance of specificity and coverage for their intended purpose, a feature that has as much to do with the design of the probe as it has for the topology of the target site. Detailed maps of the prokaryotic 16S rRNA gene and the fungal rRNA operon, the most common molecular targets for taxonomic differentiation, are shown in Fig. 2.

Technical limitations include sequence length, copy number, and other requirements unique to the method. For example, DGGE can only accommodate very short sequences to reliably destabilize and separate heterogeneous sequences, and most NGS platforms can only sequence fragments in the range of 100-600 bp, so a molecular target must be chosen wherein adequate sequence heterogeneity exists to differentiate target organisms. For QPCR, sequences should be between 100-150 bp to efficiently amplify and render accurate quantification, so primer targets must be carefully chosen to amplify all target organisms with a high degree of specificity, usually by targeting a highly variable region (e.g., for quantification of a specific genus, species, or strain) or conserved region (e.g., for total bacteria).

Molecular target copy number influences both absolutely quantitative methods (e.g., QPCR) and pseudoquantitative methods (e.g., TRFLP, NGS), as copy number, especially of rRNA operon genes, can vary widely among taxa, skewing quantitative estimates. This is particularly true of the fungal rRNA operon (including ITS), which exhibits large copy number variation among species, and is

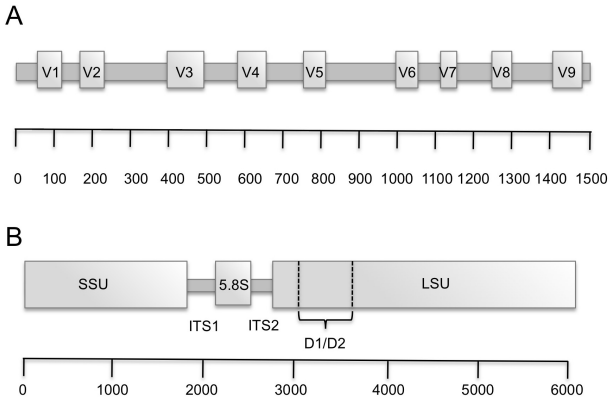

Fig. 2. Map of prokaryotic 16S rRNA gene (A) and the fungal rRNA operon (B). Scale is in bp.

not firmly quantified in most species (95). To circumvent this issue, many researchers have chosen alternative, single-copy target sites, such as the *rpoB* gene in bacteria (39), but the lack of representative sequence coverage of such genes poses a challenge for widespread implementation (see below).

Taxonomic discrimination

Community profiling methods carry the additional need for a target site that contains sufficient heterogeneity to differentiate taxa at the desired level within size constraints. For example, an optimal NGS target region should contain as much heterogeneity as possible within 100-600 bp. Not all phyla will be equally heterogeneous at a given site, however, so the "optimal" target region is dependent upon the sample type and the species contained therein (86). We have found that the V4 domain of prokaryotic 16S rDNA is a suitable region for short-amplicon NGS of food and wine fermentations, as it can reliably discriminate most LAB to genus-level in a 150 bp read (56). This is consistent with the findings of others, who bioinformatically determined the V4 domain to be appropriate for a range of environments (61), but not for all (86). Select short, hypervariable segments of the 16S rRNA gene contain a surprisingly high degree of information, rendering phylogenetic assessments comparable to those of full-length 16S sequences (61).

Fungi (and other eukarya) present a similar situation. Most yeasts contain high interspecific variability in the D1/D2 loop of LSU rDNA (96), yet this site cannot reliably discriminate species belonging to certain clades (97). The ITS region is generally hypervariable across all fungi, and thus has been adopted as the "universal barcode" for fungal identification (97), but the ITS1 and ITS2 subdomains display varying levels of intraspecific and intrageneric heterogeneity within different clades (87), and these regions can deliver dramatically different views of community diversity using short-amplicon NGS reads (98).

Coverage and specificity

Coverage describes how well a primer or probe matches all members of a given taxon along the target binding site. For example, a universal bacterial primer should ideally be homologous to all bacteria at a single, designated site, otherwise it will give an inaccurate estimate of bacterial quantity (QPCR) or community structure (profiling methods). A perfect scenario is highly unlikely, as 1) sequence heterogeneity along most universal genes dictates that at least some bacteria will have mismatches, increasing the potential for nonbinding or preferential amplification of targets displaying full homology and 2) public collections of sequence data on all microbial genes are incomplete and do not comprehend even all currently described microbial species, increasing the probability that current "universal" primers (based on sequence alignments of currently described species) do not capture this potential diversity. This is particularly pronounced for fungi, with ITS sequences in the INSD representing $< 1\%$ of known fungal species (98). For prokaryotes, a number of degenerate, high-coverage primers targeting hypervariable domains across the 16S rRNA gene have been developed and bioinformatically validated (86). For fungi, primer choice is less straightforward, as different established "universal" primers exhibit bias toward specific clades (99). Most fungal primers targeting the rRNA operon were designed 10-20 years ago based on contemporary sequence data, and a need exists for a modern revision of these primers based on current data.

Specificity describes the degree to which a primer or probe exclusively binds members of the target taxon. For example, a universal bacterial primer should not also bind archaeal or eukaryotic sequences and a primer for *Lactobacillus helveticus* should not also bind *Lactobacillus casei*. This is particularly important for QPCR, as mispriming will result in overestimation of the target, but also impacts profiling tools, as it can limit detection of target taxa. For example, many prokaryotic (100) and fungal (101) rDNA primers contain high homology to chloroplast and mitochondrial sequences. As almost all foods are plant- or animal-based, the majority of signal (e.g., DGGE, TRFLP, FISH) or sequence reads (NGS) may consist of sequences from the source material, dampening detection of the microbiota present, even if it can be unambiguously distinguished from microbial OTUs. Primers that do not bind plant/plastid DNA have been developed for both fungi (101) and bacteria (100), but these primers have not been validated against a comprehensive database and their increased specificity may risk compromising the universal coverage of these kingdoms. Thus, all primers used for molecular tools should be computationally validated to ensure adequate coverage of and specificity for target taxa. This must be balanced against the threat of eukaryotic DNA contamination on a sample-dependent basis when choosing the right target site and probe.

Selection

All factors considered, it is difficult to distinguish any one site as the optimal target for most molecular methods. For studying prokaryotic communities, the rRNA operon is historically the most common target, particularly the small-subunit 16S rRNA gene, as it contains highly conserved regions (ideal for universal primer sites) interpolated by highly variable regions (ideal for taxonomic discrimination). This gene remains the most prevalent target, as it is also the best-represented prokaryotic gene contained in all public sequence databases. For over 20 years this gene has proven its worth for phylogenetic comparison and identification of prokaryotes, but some properties of this gene make it less favorable for specific applications. Copy number variation can skew accurate quantification, making single-copy genes more appropriate for QPCR of some targets. In addition, some closely related species cannot be discriminated by short 16S rDNA amplicons, including some LAB (39). Thus, some protein-encoding genes have gained favor, such as *rpoB* (39). Such targets are poorly characterized across all bacterial phyla, however, so their degree of conservation is not well established. This increases the risk that a chosen primer/probe sequence will lack suitable specificity for and/or coverage of their intended target taxon. As

whole-genome and environmental metagenomic sequencing data continue to increase, the distribution and conservation of alternative gene sequences will become more apparent, possibly leading to the identification of a superior molecular target for microbial diversity studies.

For fungi, the situation is more complicated, and attention has fluctuated between the small-subunit 18S rRNA gene (following the lead of prokaryotic systematics) to the large subunit rRNA gene and to the ITS (102). The ITS has finally been chosen as the "universal barcode" for fungi, as it exhibits the best discrimination of most fungal species (97), but is less heterogeneous among certain fungal clades (87) and, as it is a non-coding, hypervariable region, lacks the phylogenetically metric nature of conserved rRNA genes for comparing highly divergent clades (103).

CONCLUSION

Integrating multiple molecular approaches, including high-throughput community profiling, targeted profiling, and *in situ* visualization, may provide deeper insight into food fermentations at a systems level. Statistically supported trending of microbial communities and expression profiles with fermentation outcomes (e.g., evolution of desirable sensory characteristics or of spoilage) may help identify marker genes or microbial taxa associated with specific activities. This may lead to improved strain selection for microbial inoculation and enhanced management strategies for fermentation control. Microbial ecology of these foods may no longer consider the fermentation itself, but also how the surrounding environment impacts this process. This may be viewed at both the local level—the microbial ecology of surfaces and processing equipment within the complete production facility and their relationship to the product—as well as the macro-scale—how region and climatic conditions impact the microbial succession of fermentations or raw ingredients between geographically disparate areas. Microbial community profiles generated by NGS and other emerging technologies may even be used to determine the source of foods or raw materials, including validating food products with protected-origin status (e.g., by European law). Incorporating geospatial community profiling into the systems-level framework facilitates the process of strain selection by increasing the rate of bioprospecting and marker identification, and improves opportunities for process improvement through better understanding of the relationships between substrate conditions, process parameters, microbial activity, and fermentation outcomes.

Modern molecular tools provide the opportunity for high-resolution analysis of fermentation systems, supported by the growing quality and availability of public sequence data and bioinformatic tools. NGS, in particular, is revolutionizing the way we approach microbial ecology, promising groundbreaking discoveries in the field of food microbiology. Many challenges remain, however, and researchers must carefully select the techniques, molecular targets, and analysis tools most appropriate to their experimental purpose.

REFERENCES

1. Hayes, M., Ross, R. P., Fitzgerald, G. F. and Stanton, C. (2007) Putting microbes to work: dairy fermentation, cell factories and bioactive peptides. Part I: overview. *Biotechnol. J.* **2,** 426-434.
2. Nout, M. J. (2009) Rich nutrition from the poorest - cereal fermentations in Africa and Asia. *Food Microbiol.* **26**, 685-692.
3. Ross, R. P., Morgan, S. and Hill, C. (2002) Preservation and fermentation: past, present and future. *Int. J. Food Microbiol.* **79,** 3-16.
4. Rouse, S. and van Sinderen, D. (2008) Bioprotective potential of lactic acid bacteria in malting and brewing. *J. Food Prot.* **71,** 1724-1733.
5. Ray, R. C. and Sivakumar, P. S. (2009) Traditional and novel fermented foods and beverages from tropical root and tuber crops: review. *Int. J. Food Sci. Technol.* **44**, 1073-1087.
6. Tannock, G. W. (2002) Probiotics and prebiotics: Where are we going? (eds.), Caister Academic Press, Norfork, England.
7. Heard, G. M. and Fleet, G. H. (1986) Evaluation of selective media for enumeration of yeast during wine fermentation. *J. Appl. Bacteriol.* **60,** 477-481.
8. Ampe, F., Omar, N. B., Moizan, C., Wacher, C. and Guyot, J. P. (1999) Polyphasic study of the spatial distribution of microorganisms in mexican pozol, a fermented maize dough, demonstrates the need for cultivation-independent methods to investigate traditional fermentations. *Appl. Environ. Microbiol.* **65**, 5464-5473.
9. Divol, B. and Lonvaud-Funel, A. (2005) Evidence for viable but nonculturable yeasts in botrytis-affected wine. *J. Appl. Microbiol.* **99**, 85-93.
10. Millet, V. and Lonvaud-Funel, A. (2000) The viable but non-culturable state of wine micro-organisms during storage. *Lett. Appl. Microbiol.* **30**, 136-141.
11. Bottari, B., Ercolini, D., Gatti, M. and Neviani, E. (2006) Application of FISH technology for microbiological analysis: current state and prospects. *Appl. Microbiol. Biotechnol.* **73**, 485-494.
12. Andorra, I., Monteiro, M., Esteve-Zarzoso, B., Albergaria, H. and Mas, A. (2011) Analysis and direct quantification of Saccharomyces cerevisiae and Hanseniaspora guilliermondii populations during alcoholic fermentation by fluorescence *in situ* hybridization, flow cytometry and quantitative PCR. *Food Microbiol.* **28**, 1483-1491.
13. Xufre, A., Albergaria, H., Inacio, J., Spencer-Martins, I. and Girio, F. (2006) Application of fluorescence *in situ* hybridisation (FISH) to the analysis of yeast population dynamics in winery and laboratory grape must fermentations. *Int. J. Food Microbiol.* **108**, 376-384.
14. Blasco, L., Ferrer, S. and Pardo, I. (2003) Development of specific fluorescent oligonucleotide probes for *in situ* identification of wine lactic acid bacteria. *FEMS Microbiol. Lett.* **225,** 115-123.
15. Yasuhara, T., Yuuki, T. and Kagami, N. (2001) Novel quantitative method for detection of Pectinatus using rRNA targeted fluorescent probes. *J. Am. Soc. Brew.*

Chem. **59**, 117-121.
16. Babot, J. D., Hidalgo, M., Arganaraz-Martinez, E., Apella, M. C. and Perez Chaia, A. (2011) Fluorescence *in situ* hybridization for detection of classical propionibacteria with specific 16S rRNA-targeted probes and its application to enumeration in Gruyere cheese. *Int. J. Food Microbiol.* **145**, 221-228.
17. Bottari, B., Santarelli, M., Neviani, E. and Gatti, M. (2010) Natural whey starter for Parmigiano Reggiano: culture-independent approach. *J. Appl. Microbiol.* **108**, 1676-1684.
18. Mounier, J., Monnet, C., Jacques, N., Antoinette, A. and Irlinger, F. (2009) Assessment of the microbial diversity at the surface of Livarot cheese using culture-dependent and independent approaches. *Int. J. Food Microbiol.* **133**, 31-37.
19. Ercolini, D., Hill, P. J. and Dodd, C. E. (2003) Development of a fluorescence *in situ* hybridization method for cheese using a 16S rRNA probe. *J. Microbiol. Methods* **52**, 267-271.
20. Klug, B., Rodler, C., Koller, M., Wimmer, G., Kessler, H. H., Grube, M. and Santigli, E. (2011) Oral biofilm analysis of palatal expanders by fluorescence in-situ hybridization and confocal laser scanning microscopy. *J. Vis. Exp.* **20**, 2967.
21. Postollec, F., Falentin, H., Pavan, S., Combrisson, J. and Sohier, D. (2011) Recent advances in quantitative PCR (qPCR) applications in food microbiology. *Food Microbiol.* **28**, 848-861.
22. Lucas, P. M., Claisse, O. and Lonvaud-Funel, A. (2008) High frequency of histamine-producing bacteria in the enological environment and instability of the histidine decarboxylase production phenotype. *Appl. Environ. Microbiol.* **74**, 811-817.
23. Arena, M. P., Romano, A., Capozzi, V., Beneduce, L., Ghariani, M., Grieco, F., Lucas, P. and Spano, G. (2011) Expression of Lactobacillus brevis IOEB 9809 tyrosine decarboxylase and agmatine deiminase genes in wine correlates with substrate availability. *Lett. Appl. Microbiol.* **53**, 395-402.
24. Ladero, V., Coton, M., Fernandez, M., Buron, N., Cruz Martin, M., Guichard, H., Coton, E. and Alvarez, M. A. (2011) Biogenic amines content in Spanish and French natural ciders: Application of qPCR for quantitative detection of biogenic amine-producers. *Food Microbiol.* **28**, 554-561.
25. Fernandez, M., del Rio, B., Linares, D. M., Martin, M. C. and Alvarez, M. A. (2006) Real-time polymerase chain reaction for quantitative detection of histamine-producing bacteria: use in cheese production. *J. Dairy Sci.* **89**, 3763-3769.
26. Torriani, S., Gatto, V., Sembeni, S., Tofalo, R., Suzzi, G., Belletti, N., Gardini, F. and Bover-Cid, S. (2008) Rapid detection and quantification of tyrosine decarboxylase gene (tdc) and its expression in gram-positive bacteria associated with fermented foods using PCR-based methods. *J. Food Prot.* **71**, 93-101.
27. Ibarburu, I., Aznar, R., Elizaquivel, P., Garcia-Quintans, N., Lopez, P., Munduate, A., Irastorza, A. and Duenas, M. T. (2010) A real-time PCR assay for detection and quantification of 2-branched (1,3)-β-D−glucan producing lactic acid bacteria in cider. *Int. J. Food Microbiol.* **143**, 26-31.
28. Mendes-Ferreira, A., Barbosa, C., Jimenez-Marti, E., Del Olmo, M. L. and Mendes-Faia, A. (2010) The wine yeast strain-dependent expression of genes implicated in sulfide production in response to nitrogen availability. *J. Microbiol. Biotechnol.* **20**, 1314-1321.
29. Haakensen, M. C., Butt, L., Chaban, B., Deneer, H., Ziola, B. and Dowgiert, T. (2007) horA-Speciric real-time PCR for detection of beer-spoilage lactic acid bacteria. *J. Am. Soc. Brew. Chem.* **65**, 157-165.
30. Cho, G. S., Krauss, S., Huch, M., Du Toit, M. and Franz, C. M. (2011) Development of a quantitative PCR for detection of Lactobacillus plantarum starters during wine malolactic fermentation. *J. Microbiol. Biotechnol.* **21**, 1280-1286.
31. Muyzer, G., Dewaal, E. C. and Uitterlinden, A. G. (1993) Profiling of complex microbial populations by denaturing gradient gel-electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S ribosomal RNA. *Appl. Environ. Microbiol.* **59**, 695-700.
32. Kisand, V. and Wikner, J. (2003) Limited resolution of 16S rDNA DGGE caused by melting properties and closely related DNA sequences. *J. Microbiol. Meth.* **54**, 183-191.
33. Sekiguchi, H., Tomioka, N., Nakahara, T. and Uchiyama, H. (2001) A single band does not always represent single bacterial strains in denaturing gradient gel electrophoresis analysis. *Biotechnol. Lett.* **23,** 1205-1208.
34. Nubel, U., Engelen, B., Felske, A., Snaidr, J., Wieshuber, A., Amann, R. I., Ludwig, W. and Backhaus, H. (1996) Sequence heterogeneities of genes encoding 16S rRNAs in Paenibacillus polymyxa detected by temperature gradient gel electrophoresis. *J. Bacteriol.* **178**, 5636-5643.
35. Polz, M. F. and Cavanaugh, C. M. (1998) Bias in template-to-product ratios in multitemplate PCR. *Appl. Environ. Microbiol.* **64**, 3724-3730.
36. Speksnijder, A., Kowalchuk, G. A., De Jong, S., Kline, E., Stephen, J. R. and Laanbroek, H. J. (2001) Microvariation artifacts introduced by PCR and cloning of closely related 16S rRNA gene sequences. *Appl. Environ. Microbiol.* **67**, 469-472.
37. Tourlomousis, P., Kemsley, E. K., Ridgway, K. P., Toscano, M. J., Humphrey, T. J. and Narbad, A. (2010) PCR-denaturing gradient gel electrophoresis of complex microbial communities: a two-step approach to address the effect of gel-to-gel variation and allow valid comparisons across a large dataset. *Microb. Ecol.* **59**, 776-786.
38. Cocolin, L., Bisson, L. F. and Mills, D. A. (2000) Direct profiling of the yeast dynamics in wine fermentations. *FEMS Microbiol. Lett.* **189**, 81-87.
39. Renouf, V., Claisse, O., Miot-Sertier, C. and Lonvaud-Funel, A. (2006) Lactic acid bacteria evolution during winemaking: use of rpoB gene as a target for PCR-DGGE analysis. *Food Microbiol.* **23**, 136-145.
40. Ercolini, D. (2004) PCR-DGGE fingerprinting: novel strategies for detection of microbes in food. *J. Microbiol. Meth.* **56**, 297-314.
41. Kaplan, C. W. and Kitts, C. L. (2003) Variation between observed and true Terminal Restriction Fragment length is dependent on true TRF length and purine content. *J. Microbiol. Meth.* **54**, 121-125.

42. Marsh, T. L. (2005) Culture-independent microbial community analysis with terminal restriction fragment length polymorphism. *Methods Enzymol.* **397**, 308-329.
43. Hartmann, M., Enkerli, J. and Widmer, F. (2007) Residual polymerase activity-induced bias in terminal restriction fragment length polymorphism analysis. *Env. Microbiol.* **9**, 555-559.
44. Egert, M. and Friedrich, M. W. (2003) Formation of pseudo-terminal restriction fragments, a PCR-related bias affecting terminal restriction fragment length polymorphism analysis of microbial community structure. *Appl. Environ. Microbiol.* **69**, 2555-2562.
45. Bokulich, N. A. and Mills, D. A. (2012) Differentiation of mixed lactic acid bacteria communities in beverage fermentations using targeted terminal restriction fragment length polymorphism. *Food Microbiol.* **31**, 126-132. doi:10.1016/j.fm.2012.02.007.
46. Culman, S. W., Gauch, H. G., Blackwood, C. B. and Thies, J. E. (2008) Analysis of T-RFLP data using analysis of variance and ordination methods: A comparative study. *J. Microbiol. Meth.* **75**, 55-63.
47. Blackwood, C. B., Marsh, T. L., Kim, S. H. and Paul, E. A. (2003) Terminal restriction fragment length polymorphism data analysis for quantitative comparison of microbial communities. *Appl. Environ. Microbiol.* **69**, 926-932.
48. Blackwood, C. B., Hudleston, D., Zak, D. R. and Buyer, J. S. (2007) Interpreting ecological diversity indices applied to terminal restriction fragment length polymorphism data: insights from simulated microbial communities. *Appl. Environ. Microbiol.* **73**, 5276-5283.
49. Schutte, U. M. E., Abdo, Z., Bent, S. J., Shyu, C., Williams, C. J., Pierson, J. D. and Forney, L. J. (2008) Advances in the use of terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes to characterize microbial communities. *Appl. Microbiol. Biotechnol.* **80**, 365-380.
50. Liu, W. T., Marsh, T. L., Cheng, H. and Forney, L. J. (1997) Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl. Environ. Microbiol.* **63**, 4516-4522.
51. Rademaker, J. L. W., Hoolwerf, J. D., Wagendorp, A. A. and te Giffel, M. C. (2006) Assessment of microbial population dynamics during yoghurt and hard cheese fermentation and ripening by DNA population fingerprinting. *Int. Dairy J.* **16**, 457-466.
52. Rademaker, J. L. W., Peinhopf, M., Rijnen, L., Bockelmann, W. and Noordman, W. H. (2005) The surface microflora dynamics of bacterial smear-ripened Tilsit cheese determined by T-RFLP DNA population fingerprint analysis. *Int. Dairy J.* **15**, 785-794.
53. Sanchez, J. I., Rossetti, L., Martinez, B., Rodriguez, A. and Giraffa, G. (2006) Application of reverse transcriptase PCR-based T-RFLP to perform semi-quantitative analysis of metabolically active bacteria in dairy fermentations. *J. Microbiol. Meth.* **65,** 268-277.
54. Bokulich, N. A., Hwang, C. F., Liu, S., Boundy-Mills, K. and Mills, D. A. (2012) Profiling the yeast communities of wine using terminal restriction fragment length polymorphism. *Am. J. Enol. Vitic.* **63**, doi:10.5344/ajev.2011.11077.
55. Bokulich, N. A., Bamforth, C. W. and Mills, D. A. (2012) Brewhouse-resident microbiota are responsible for multi-stage fermentation of american coolship ale. *PLoS ONE* **7**, e35507. doi:10.1371/journal.pone.0035507.
56. Bokulich, N. A., Joseph, C. M. L., Allen, G., Benson, A. K. and Mills, D. A. (2012) Next-Generation Sequencing Reveals Significant Bacterial Diversity of Botrytized Wine. *PLoS ONE* **7**, e36357. doi:10.1371/journal.pone.0036357.
57. Marcobal, A., Underwood, M. and Mills, D. A. (2008) Rapid determination of the bacterial composition of commercial probiotic products by terminal restriction fragment length polymorphism analysis. *J. Pediatr. Gastroenterol. Nutr.* **46**, 608-611.
58. Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L., Jarvie, T. P., Jirage, K. B., Kim, J. B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F. and Rothberg, J. M. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-380.
59. Bennett, S. (2004) Solexa Ltd. *Pharmacogenomics* **5**, 433-438.
60. Kuczynski, J., Lauber, C. L., Walters, W. A., Parfrey, L. W., Clemente, J. C., Gevers, D. and Knight, R. (2012) Experimental and analytical tools for studying the human microbiome. *Nat. Rev. Genet.* **13**, 47-58.
61. Liu, Z., Lozupone, C. A., Hamady, M., Bushman, F. D. and Knight, R. (2007) Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids. Res.* **35**, e120.
62. Kiyohara, M., Koyanagi, T., Matsui, H., Yamamoto, K., Take, H., Katsuyama, Y., Tsuji, A., Miyamae, H., Kondo, T., Nakamura, S., Katayama, T. and Kumagai, H. (2012) Changes in microbiota population during fermentation of narezushi as revealed by pyrosequencing analysis. *Biosci. Biotechnol. Biochem.* **76**, 48-52.
63. Koyanagi, T., Kiyohara, M., Matsui, H., Yamamoto, K., Kondo, T., Katayama, T. and Kumagai, H. (2011) Pyrosequencing survey of the microbial diversity of 'narezushi', an archetype of modern Japanese sushi. *Lett. Appl. Microbiol.* **53**, 635-640.
64. Sakamoto, N., Tanaka, S., Sonomoto, K. and Nakayama, J. (2011) 16S rRNA pyrosequencing-based investigation of the bacterial community in nukadoko, a pickling bed of fermented rice bran. *Int. J. Food Microbiol.* **144**, 352-359.
65. Li, X. R., Ma, E. B., Yan, L. Z., Meng, H., Du, X. W., Zhang, S. W. and Quan, Z. X. (2011) Bacterial and fungal diversity in the traditional Chinese liquor fermentation process. *Int. J. Food Microbiol.* **146**, 31-37.
66. Alegria, A., Szczesny, P., Mayo, B., Bardowski, J. and Kowalczyk, M. (2012) Biodiversity in Oscypek, a traditional Polish cheese, determined by culture-dependent

and -independent approaches. *Appl. Environ. Microbiol.* **78**, 1890-1898.
67. Humblot, C. and Guyot, J. P. (2009) Pyrosequencing of tagged 16S rRNA gene amplicons for rapid deciphering of the microbiomes of fermented foods such as pearl millet slurries. *Appl. Environ. Microbiol.* **75**, 4354-4361.
68. Roh, S. W., Kim, K. H., Nam, Y. D., Chang, H. W., Park, E. J. and Bae, J. W. (2010) Investigation of archaeal and bacterial diversity in fermented seafood using barcoded pyrosequencing. *ISME J.* **4**, 1-16.
69. Jung, M. J., Nam, Y. D., Roh, S. W. and Bae, J. W. (2012) Unexpected convergence of fungal and bacterial communities during fermentation of traditional Korean alcoholic beverages inoculated with various natural starters. *Food Microbiol.* **30**, 112-123.
70. Kim, Y. S., Kim, M. C., Kwon, S. W., Kim, S. J., Park, I. C., Ka, J. O. and Weon, H. Y. (2011) Analyses of bacterial communities in meju, a Korean traditional fermented soybean bricks, by cultivation-based and pyrosequencing methods. *J. Microbiol.* **49**, 340-348.
71. Nam, Y. D., Lee, S. Y. and Lim, S. I. (2012) Microbial community analysis of Korean soybean pastes by next-generation sequencing. *Int. J. Food Microbiol.* **155**, 36-42.
72. Nam, Y. D., Park, S. L. and Lim, S. I. (2012) Microbial Composition of the Korean Traditional Food "kochujang" Analyzed by a Massive Sequencing Technique. *J. Food Sci.* **77**, M250-256.
73. Park, E. J., Chun, J., Cha, C. J., Park, W. S., Jeon, C. O. and Bae, J. W. (2012) Bacterial community analysis during fermentation of ten representative kinds of kimchi with barcoded pyrosequencing. *Food Microbiol.* **30**, 197-204.
74. Park, E. J., Kim, K. H., Abell, G. C., Kim, M. S., Roh, S. W. and Bae, J. W. (2011) Metagenomic analysis of the viral communities in fermented foods. *Appl. Environ Microbiol.* **77**, 1284-1291.
75. Lozupone, C. A. and Knight, R. (2005) UniFrac: A new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71**, 8228-8235.
76. Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Gonzalez Pena, A., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J. and Knight, R. (2010) Qiime allows analysis of high-throughput community sequence data. *Nat. Methods* **7**, 335-336.
77. Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J. and Weber, C. F. (2009) Introducing mothur: open-source, platform- independent, community-supported software for describing and comparing microbial communities. *Appl. Environ Microbiol.* **75**, 7537-7541.
78. Zaneveld, J. R., Parfrey, L. W., Van Treuren, W., Lozupone, C., Clemente, J. C., Knights, D., Stombaugh, J., Kuczynski, J. and Knight, R. (2011) Combined phylogenetic and genomic approaches for the high-throughput study of microbial habitat adaptation. *Trends Microbiol.* **19**, 472-482.
79. Nilsson, R. H., Ryberg, M., Kristiansson, E., Abarenkov, K., Larsson, K. H. and Koljalg, U. (2006) Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. *PLoS ONE* **1**, e59.
80. Tedersoo, L., Abarenkov, K., Nilsson, R. H., Schussler, A., Grelet, G. A., Kohout, P., Oja, J., Bonito, G. M., Veldre, V., Jairus, T., Ryberg, M., Larsson, K. H. and Koljalg, U. (2011) Tidying up international nucleotide sequence databases: ecological, geographical and sequence quality annotation of its sequences of mycorrhizal fungi. *PLoS ONE* **6**, e24940.
81. McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., DeSantis, T. Z., Probst, A., Andersen, G. L., Knight, R. and Hugenholtz, P. (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **6**, 610-618.
82. De Santis, T., Hugenholtz, P., Larsen, N., Rojas, N., Brodie, E., Keller, K., Huber, T., Dalevi, D., Hu, P. and Andersen, G. L. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**, 5069-5072.
83. Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J. and Glockner, F. O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids. Res.* **35**, 7188-7196.
84. Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., Kulam-Syed-Mohideen, A. S., McGarrel, D. M., Marsh, T. L., Garrity, G. M. and Tiedje, J. M. (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids. Res.* **37,** D141-D145.
85. Abarenkov, K., Nilsson, R. H., Larsson, K.-H., Alexander, I. J., Eberhardt, U., Erland, S., Høiland, K., Kjøller, R., Larsson, E., Pennanen, T., Sen, R., Taylor, A. F. S., Tedersoo, L., Ursing, B. M., Vrålstad, T., Liimatainen, K., Peintner, U. and Kõljalg, U. (2010) The UNITE database for molecular identification of fungi - recent updates and future perspectives. *New Phytol.* **186**, 1447-1452.
86. Soergel, D. A. W., Dey, N., Knight, R. and Brenner, S. E. (2012) Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. *ISME J.* **7**, doi: 10.1038/ismej.2011.208.
87. Nilsson, R. H., Kristiansson, E., Ryberg, M., Hallenberg, N. and Larsson, K. H. (2008) Intraspecific ITS variability in the kingdom fungi as expressed in the international sequence databases and its implications for molecular species identification. *Evol. Bioinform. Online* **4**, 193-201.
88. Simon, C. and Daniel, R. (2011) Metagenomic analyses: past and future trends. *Appl. Environ. Microbiol.* **77**, 1153-1161.
89. Lauber, C. L., Zhou, N., Gordon, J. I., Knight, R. and Fierer, N. (2010) Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. *FEMS Microbiol. Lett.* **307**, 80-86.
90. Li, F., Hullar, M. A. and Lampe, J. W. (2007) Optimization of terminal restriction fragment polymorphism (TRFLP) analysis of human gut microbiota. *J. Microbiol. Meth.* **68**, 303-311.

91. Rossen, L., Norskov, P., Holmstrom, K. and Rasmussen, O. F. (1992) Inhibition of PCR by components of food samples, microbial diagnostic assays and DNA-extraction solutions. *Int. J. Food Microbiol.* **17**, 37-45.
92. Nocker, A., Cheung, C. Y. and Camper, A. K. (2006) Comparison of propidium monoazide with ethidium monoazide for differentiation of live vs. dead bacteria by selective removal of DNA from dead cells. *J. Microbiol. Meth.* **67**, 310-320.
93. Nocker, A., Richter-Heitmann, T., Montijn, R., Schuren, F. and Kort, R. (2010) Discrimination between live and dead cells in bacterial communities from environmental water samples analyzed by 454 pyrosequencing. *Int. Microbiol.* **13**, 59-65.
94. Andorra, I., Esteve-Zarzoso, B., Guillamon, J. M. and Mas, A. (2010) Determination of viable wine yeast using DNA binding dyes and quantitative PCR. *Int. J. Food Microbiol.* **144**, 257-262.
95. Anderson, I. C. and Cairney, J. W. G. (2004) Diversity and ecology of soil fungal communities: increased understanding through the application of molecular techniques. *Env. Microbiol.* **6**, 769-779.
96. Kurtzman, C. and Robnett, C. J. (1998) Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. *Antonie van Leeuwenhoek* **73**, 331-371.
97. Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A. and Chen, W. (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *PNAS* **109**, 6241-6246.
98. Nilsson, R. H., Ryberg, M., Abarenkov, K., Sjokvist, E. and Kristiansson, E. (2009) The ITS region as a target for characterization of fungal communities using emerging sequencing technologies. *FEMS Microbiol. Lett.* **296**, 97-101.
99. Bellemain, E., Carlsen, T., Brochmann, C., Coissac, E., Taberlet, P. and Kauserud, H. (2010) ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. *BMC Microbiol.* **10**, 189.
100. Sakai, M., Matsuka, A., Komura, T. and Kanazawa, S. (2004) Application of a new PCR primer for terminal restriction fragment length polymorphism analysis of the bacterial communities in plant roots. *J. Microbiol. Methods.* **59**, 81-89.
101. Martin, K. J. and Rygiewicz, P. T. (2005) Fungal-specific PCR primers developed for analysis of the ITS region of environmental DNA extracts. *BMC Microbiol.* **5**, 28.
102. Sakamoto, N., Tanaka, S., Sonomoto, K. and Nakayama, J. (2011) 16S rRNA pyrosequencing-based investigation of the bacterial community in nukadoko, a pickling bed of fermented rice bran. *Int. J. Food Microbiol.* **144**, 352-359.
103. Kruger, D., Kapturska, D., Fischer, C., Daniel, R. and Wubet, T. (2012) Diversity Measures in Environmental Sequences Are Highly Dependent on Alignment Quality-Data from ITS and New LSU Primers Targeting Basidiomycetes. *PLoS ONE* **7**, e32139.