Published in final edited form as:

J Microbiol Methods. 2012 September ; 90(3): 235–240. doi:10.1016/j.mimet.2012.05.013.

Improving oligonucleotide fingerprinting of rRNA genes by implementation of polony microarray technology

Paul M. Ruegger^a, Elizabeth Bent^a, Wei Li^c, Daniel R. Jeske^b, Xinping Cui^b, Jonathan Braun^d, Tao Jiang^c, and James Borneman^{a,*}

^aDepartment of Plant Pathology and Microbiology, University of California, Riverside, CA, USA

^bDepartment of Statistics, University of California, Riverside, CA, USA

^cDepartment of Computer Science, University of California, Riverside, CA, USA

^dDepartment of Molecular and Medical Pharmacology, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

Abstract

Improvements to oligonucleotide fingerprinting of rRNA genes (OFRG) were obtained by implementing polony microarray technology. OFRG is an array-based method for analyzing microbial community composition. Polonies are discrete clusters of DNA, produced by solid-phase PCR in hydrogels, and derived from individual, spatially isolated DNA molecules. The advantages of a polony-based OFRG method include higher throughput and reductions in the PCR-induced errors and compositional skew inherent in all other PCR-based community composition methods, including high throughput sequencing of rRNA genes. Given the similarities between polony microarrays and certain aspects of sequencing methods such as the Illumina platform, we suggest that if concepts presented in this study were implemented in high throughput sequencing protocols, a reduction of PCR-induced errors and compositional skew may be realized.

Keywords

Oligonucleotide fingerprinting of rRNA genes; OFRG; Polony; Polonies; PCR-induced errors; High throughput sequencing

1. Introduction

Methods for studying microbial community composition range from inexpensive, coarse-grained approaches such as denaturing gradient gel electrophoresis (DGGE) (Muyzer et al., 1993), terminal restriction fragment length polymorphism (T-RFLP) (Liu et al., 1997) and ribosomal intergenic spacer analysis (RISA) (Borneman and Triplett, 1997), to the significantly more expensive and more accurate “gold standard” of sequencing near full-length rRNA genes (Barns et al. 1994). Sequencing near full-length 16S rRNA genes

© 2012 Elsevier B.V. All rights reserved

*Corresponding author at: Department of Plant Pathology and Microbiology, University of California, Riverside, CA, USA. Tel.: 1 951 8273584; fax: 1 951 8274294. borneman@ucr.edu (J. Borneman)..

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

provides the highest available taxonomic resolution, and it can be reasonably affordable when an accurate “snapshot” of a microbial community is required. However, though costs are dropping, multi-sample longitudinal studies that employ in-depth full-length sequencing are often still too expensive for many labs. Recent developments in the use of high-throughput sequencing enable cost-effective in-depth analysis, but due to the short read-lengths, these methods are limited in their taxonomic resolution (Caporaso et al., 2011; Price et al. 2009; Wu et al., 2010).

The focus of this research was on improving an alternative method termed oligonucleotide fingerprinting of ribosomal rRNA genes (OFRG) (Borneman et al., 2001; Valinsky et al., 2002). The OFRG method uses a set of 40 computer-designed DNA probes chosen from a set of training sequences, and hybridizes them against an array of sample-derived microbial rRNA gene clones. The hybridization affinity of each probe/clone combination is quantified and then processed in one of two ways. Traditionally these data are transformed into a 40-digit binary “fingerprint” for each clone, where a 0 denotes no hybridization and a 1 denotes a successful hybridization event (Figuerola et al. 2004; Yu et al., 2010). These fingerprints are then clustered based on their similarity to the fingerprints of other clones in the array. Alternatively, hierarchical clustering can be performed on the data without a binary transformation of the hybridization intensities. In either case, these clusters provide a low-cost estimate of the relative proportions of the various microbial taxa present in an environment because similar fingerprints arise from similar rRNA gene sequences. Clusters can be identified by their relationships to known rRNA genes in the cluster analysis. An important feature of OFRG is that it utilizes near full-length rRNA genes, often enabling species-level resolution (Valinsky et al., 2002) (Ruegger et al, 2011).

To improve the OFRG method, we implemented polony microarray technology (Mitra and Church, 1999). Polonies, or “polymerase colonies,” are analogous to the spots on an array. Rather than being printed, however, polonies are grown in place; each polony consists of many thousands of localized copies of an individual DNA molecule generated through solid-phase PCR in a polyacrylamide hydrogel. Because diffusion of the PCR amplicons is inhibited somewhat by the gel, the process results in the formation of spots of DNA molecules – the polonies – randomly placed in the gel, one for each original DNA molecule. Polony technology has been employed in several applications such as DNA sequencing, SNP detection, gene expression studies, genotyping, haplotyping and alternative pre-mRNA splicing (Butz et al. 2004; Mitra et al. 2003; Rieger et al. 2007; Shendure et al. 2005; Zhang et al., 2006; Zhu et al. 2003).

Three key characteristics of polony microarrays enable improvements to the OFRG method. These include 1) spatial isolation of DNA template before PCR amplification, 2) polony generation directly from genomic DNA template, and 3) increased throughput. Though the latter is clearly beneficial, it may not be readily apparent how the first two characteristics enable improvements to OFRG, or how high-throughput sequencing methods might also benefit from their implementation. Many microbial community analysis strategies such as those listed above necessarily employ a mixed-template PCR step. For instance, an initial mixed-template PCR is required in high-throughput sequencing to prepare DNA templates suitable for a second PCR that generates clusters on the surface of a flow cell. During a mixed-template PCR, errors such as chimera formation (Qiu et al., 2001; Wang and Wang, 1996) and population skew (Chandler et al. 1997; Farrelly et al., 1995; Reysenbach et al. 1992; Suzuki and Giovannoni, 1996) may be introduced and propagate. Chimeras form in conventional reaction tubes during mixed-template PCRs because amplicons of all species can interact freely; during the annealing step, inter-species hybridizations of partially extended amplicons can be fully extended in a subsequent cycle, thus creating chimeras. In contrast, genomic DNA template in a polony microarray is spatially isolated before any PCR

amplification occurs, essentially transforming a mixed-template PCR into multitudes of single-template PCRs. Thus, the polony microarray affords a simple and elegant strategy to limit interaction between heterogeneous amplicons and reduce the formation of chimeric amplicons. Population skew also occurs in mixed-template, tube-based reactions due to the phenomenon of PCR bias; the DNA of some species amplifies more efficiently than others, respectively leading to over- or underrepresentation in the PCR product as compared to their true population levels in the original sample. Such amplification bias is still present in the PCR used to create a polony microarray. However, polonies are created with a single PCR reaction that starts with (spatially separated) genomic DNA as template. Because of these unique conditions, rather than skewing the final proportions of bacterial species, the effect of PCR bias is manifested as a skewing of the amount of DNA produced in each polony; DNA from easily amplified species results in larger polonies and DNA from difficult to amplify species results in smaller polonies. Lastly, increased throughput, from 9,600-clone macroarrays (Bent et al., 2006) to microarrays containing millions of polonies (Mitra and Church, 1999) are possible.

2. Materials and methods

2.1. Experimental setup

To demonstrate the capabilities of polony OFRG we performed analyses on known and unknown DNA samples. To highlight an unusual characteristic of polony technology, having potential benefits for high-throughput microbial community analyses, we compared results of pre-PCR and no pre-PCR treatments on the sample DNA.

For the known bacterial community experiment, our primary aim was to determine the ability of polony OFRG to distinguish bacterial 16S rRNA genes at various taxonomic levels. To do so, we chose ten bacterial clones from a clone library, some being closely related and others being more distantly related to each other. A secondary aim was to highlight the potential benefit of polony OFRG to reduce PCR artifacts by spatially isolating genomic DNA prior to downstream processing and analysis. To achieve this, we compared results from polonies made with the above ten bacterial clones as starting template (akin to genomic DNA) to results from polonies made with the 35-cycle PCR products of those same clones as starting template (akin to the starting template used in high-throughput sequencing, as well as our previous OFRG protocol). A tertiary aim was to assess the repeatability of polony OFRG, so we performed three technical replicates for each starting template.

For the multi-environment experiment, our primary aim was to determine whether a polony OFRG analysis could distinguish whole microbial communities of unknown composition. The focus of our research is inflammatory bowel disease and so we chose to analyze samples from healthy, ulcerative colitis and Crohn's disease patients. However, our knowledge of the microbial populations in these samples was limited to only a few phylotypes (Presley et al, 2012) and we were unsure whether and to what extent community-wide differences could be detected. We therefore included an ocean sample in the experiment to help ensure some contrast in the results, though any nongut sample would have sufficed. Finally, as an additional test of repeatability, we performed three technical replicates of each starting template in this experiment as well.

2.2. DNA template selection for 10-clone experiment

Near full-length 16S rRNA gene sequences from our clone library, created during previous experiments from mouse intestinal samples, were classified using the Ribosomal Database Project's Naïve Bayesian Classifier (Wang et al., 2007). After classification, ten sequences were chosen from eight genera and six phyla. Their classifications are shown in an

abbreviated format (Clone_ID.Phylum;Class;Genus, with classification confidence values after each level) in Figure 1. The dendrogram in Figure 1 was created within Geneious, version 5.1 (Drummond et al., 2010), by first performing an alignment of the sequences using Muscle (Edgar, 2004) with default parameters. The UPGMA tree was created from the alignment using Geneious Tree Builder's default parameters.

2.3. DNA template preparation

For the artificial community experiment, ten 16S rRNA genes were cloned into the pNEB205A plasmid (New England Biolabs, Ipswich, MA). Plasmid concentrations were determined with a NanoDrop spectrophotometer (NanoDrop Products, Wilmington, DE), mixed in equimolar amounts and diluted to 0.3 pg/ μ l.

Using the mixture of plasmids as template, the 35-cycle template was prepared by performing four 10 μ l amplification reactions in a MJ Research PTC-200 thermal cycler (Bio-Rad Laboratories Inc., Hercules, CA) containing the following reagents: 50 mM Tris (pH 8.3), 500 μ g/ml bovine serum albumin (BSA), 2.5 mM $MgCl_2$, 250 μ M each deoxynucleotide triphosphate (dNTP), 400 nM each primer, and 0.5 units *Taq* DNA polymerase. Thermal cycling parameters were: 94 °C for 2 min; 35 cycles of 94 °C 20 s, 50 °C 30 s, 72 °C 60 s; 72 °C 4 min. The PCR primers were M13F (TGTAACGACGCGCCAGT) and M13R (GAAACAGCTATGACCATGAT). The four replicate amplification reactions were pooled and gel purified using the QIAquick Gel Extraction Kit (Qiagen, Valencia, CA) without the use of ethidium bromide staining or UV light. The purified template was diluted 10⁵-fold for use. Colonies were made with either the 0-PCR-cycle plasmid DNA or the purified 35-cycle DNA dilutions. The optimal amount of template to add was determined empirically using polony microarrays. A total of 1.6 μ l of the plasmid mixture was added to each 0-cycle gel and 0.4 μ l of the 35-cycle DNA was added to each 35-cycle gel.

For the multi-environment experiment, three intestinal samples and one ocean sample were selected. The ocean DNA came from one gallon of Pacific Ocean water (collected in Oceanside, CA, latitude 33.193104, longitude -117.386267), which was first passed through a 0.22 μ m filter (Millipore Co., Billerica, MA). The filter membrane was removed, air-dried and cut into small pieces. DNA was extracted from the filter membrane using the FastDNA Spin Kit (Qbiogene, Carlsbad, CA) as described by the manufacturer. DNA was further purified by subjecting it to electrophoresis on 1% agarose gel and then isolating the DNA above 3 Kb using the QIAquick Gel Extraction Kit (Qiagen, Valencia, CA), without exposing the DNA to UV or ethidium bromide. The DNA was resuspended in 62 μ l of 1/10th EB and the concentration was determined to be 102.5 ng/ μ l using a NanoDrop spectrophotometer. The ocean water DNA was diluted 10³-fold in 1/10th EB for use. The intestinal DNA used in this study was described previously in Presley et al. (2012). Briefly, mucosal luminal interface samples were obtained using an endoscopic saline-lavage sampling technique. Samples (~30 ml) were centrifuged for 1 minute at 14,000 $\times$ g. Pellets were resuspended with 1 ml CLS-Y buffer and transferred to Lysis Tubes from a FastDNA Spin Kit (Qbiogene, Carlsbad, CA). DNA was extracted using the FastDNA Spin Kit (Qbiogene, Carlsbad, CA) as described by the manufacturer, with a 30-second bead-beating step at a FastPrep instrument setting of 5.5. DNA was further purified by subjecting it to electrophoresis in a 1% agarose gel and then recovering it from the gel using a QIAquick Gel Extraction Kit (Qiagen, Valencia, CA) as described by the manufacturer, but without exposing the DNA to UV or ethidium bromide and excluding the heating step to dissolve the gel in Buffer QG. The DNA was resuspended in 50 μ l of 1/10th EB, vacuum dried and stored at -80 °C. For use in polony experiments, the DNA was rehydrated to its original volume with ultrapure water then diluted 30-fold with 1/10th EB. The optimal volumes of each template to add from each environment were determined empirically using polony

microarrays, and were, 0.70 μ l for CD, 2.50 μ l for UC, 1.00 μ l for healthy control, and 2.40 μ l for the ocean sample.

2.4. Polony slide preparation

Teflon-coated slides (17 $\times$ 40 mm, single-well, Thermo Fisher, Waltham, MA) were cleaned by submersing them in 2 M HCl for 2 hours with gentle stirring, then triple-rinsing in ultrapure water. The slides were then treated with Bind Silane (GE Healthcare Biosciences, Pittsburgh, PA) by submersing them in a treating solution (193 ml 100% EtOH, 6 ml ultrapure water, 600 μ l 17 M glacial acetic acid, and 500 μ l Bind Silane) overnight. Slides were washed 2 times in 100% EtOH, allowed to air dry, heated at 60 °C for 2 hours, allowed to cool, and then stored in a desiccator.

2.5. Polony microarray production

Polony microarrays were created by adding DNA template to an acrylamide gel mixture, casting the mixture onto a prepared microscope slide, and then performing solid-phase PCR in the cured gel. In the following, we use the terms slides, gels or microarrays interchangeably, depending on context.

The acrylamide gels were prepared first, and then a PCR master mix was applied to the cured gel for thermal cycling. Eighteen μ l of the gel mixture [2 μ M reverse acrydite primer (see below), 20% ABD mix (ABD mix = 450 μ l 40% acrylamide, 50 μ l 19:1 acrylamide:bis-acrylamide and 10 mg (+)-N,N'-Diallyltartramide), 2 mg/ml bovine serum albumin, 0.1% N,N,N',N'-Tetramethylethylenediamine, 0.1% ammonium persulfate] was mixed with the template and placed on a coverglass (22 $\times$ 40 mm, 0.17–0.25 mm thick). A Bind Silane-treated 17 $\times$ 40 mm, single-well, Teflon-coated slide (see above) was carefully brought upside down to the gel-mixture droplet until capillary action pulled the coverglass onto the slide; this helps prevent the formation of bubbles in the gel. The slides were allowed to polymerize for 45 minutes at RT. Coverglasses were removed and the slides were washed in ultrapure water for 15 minutes to remove unbound acrylamide. The gels were dried in air completely before adding the PCR master mix.

Forty-five μ l of PCR master mix (50 mM Tris (pH 8.3), 2.5 mM MgCl₂, 250 μ M each dNTP, 400 nM forward primer, 0.24 units/ μ l ThermoSequenase DNA polymerase (USB Corporation, Cleveland, OH) and 2 mg/ml bovine serum albumin) was spread onto the gel and allowed to soak in for at least 2 minutes. To increase the efficiency of solid-phase PCR on full-length bacterial 16S sequences, we used a modified 27F primer, adding an M13R sequence to the 5' end (ACAGGAAACAGCTATGACCATGAGRRTTTGATYHTGGYTCAG). The reverse primer is the universal bacterial primer 1492R but with an acrydite moiety at the 5' end that covalently attaches to the acrylamide gel during polymerization (reverse acrydite primer, GBTACCTTGTTACGACTT). A 20 $\times$ 35 mm coverglass was carefully lowered onto the mix to allow bubbles to escape. A 22 $\times$ 40 mm Secure-Seal hybridization chamber (SA-500, Grace Bio Labs, Bend, OR) was affixed over the coverglass and filled with mineral oil to prevent evaporation. The slides were thermal cycled in an MJ Research PTC-200 with a Twin Tower block (Bio-Rad Laboratories) using the parameters: 94 °C for 2 min; 8 cycles of 95 °C 10 s, 48 °C 30 s, 72 °C 4 min; 62 cycles of 94 °C 20 s, 50 °C 30 s, 72 °C 2 min; 72 °C 5 min. The high number of cycles helps overcome the low efficiency of solid-phase PCR. After thermal cycling, the slides were placed in a glass hexane-filled Copeland jar for 10 minutes to loosen the Secure-Seal chambers and dissolve the mineral oil. The slides were separated from the chambers and placed in another Copeland jar with fresh hexane for a few moments to remove residual mineral oil. The coverglasses were then removed and four slides each were placed into a LockMailer™ Microscope Slide Jar (Simport, Beloeil,

Quebec, Canada) for subsequent processing. To reduce background signal, 7 ml of a Proteinase K solution [20 mM Tris-HCl pH 8.0, 0.10 mg/ml Proteinase K (New England Biolabs)] was added to the jars and incubated at 37 °C for 1 hour, then washed 5 minutes in ultrapure water. To denature the double-stranded amplification products that make up the colonies, 7 ml of 70% formamide (in water) were added to each LockMailer™ Jar, and the jars were placed in a water bath at 75 °C for 15 minutes. To cool the solution, the jars were placed in RT water for 5 minutes. The formamide solution was decanted and replaced with ultrapure water and gently shaken for 3 minutes. This was repeated two more times to remove residual formamide.

2.6. Polony hybridization probes

Probes were designed computationally using the Maximum Fidelity Probe Set algorithm (Ruegger et al. 2011). To increase annealing temperatures of the short (10-mer) probes, three locked nucleic acid nucleotides were added on the 5' ends (Koshkin et al. 1998; Obika et al. 1998). Two 18-carbon spacers and a fluorophore were also added at the 5' ends. Fluorophores were either TYE™ 563 or TYE™ 665. The spacers, placed between the fluorophore and probe, were added to improve binding specificity. The probes used in this experiment having a TYE™ 563 fluorophore were: TACTGACGCT, GTTGCCAGCA, TACAATGGCG, TGAATACGTT, CAAACAGGAT, GCCGTAAACG, CAGCTAACGC, CCTGGGGAGT, CGCAACCCTT, GAGACTGCCG, TTCGGATTGT, TCAGCATGCC, AGCGTGGGGA, AAGGAATTGA, ACCTTACCTG, CTTGACATCC, TGGGGATGAC, GTCTGCAACT, TAGTAATCGC, GAGGTGGAGC. Probes having a TYE™ 665 fluorophore were: TGCAAGTCGA, ATACCGCATA, GTGAAAGCCC, GCGAAGGCGG, GCCTAACACA, GCTAGTTGGT, GCTCACCAAG, ACCAAGGCCG, ACAATGGGCG, CAATGGGCGA, GATGAAGGCC, GACGCTGAGG, CTAATACCGG, GGTGGGGTAA, GCCCACCAAG, GACTGAGACA, CCAGACTCCT, TGCAAGCGTT, GGGCTCAACC, GCAAGTCGAA. Note that these probe types were hybridized in ordered pairs, one of each fluorophore type, as listed above. In addition, reference probes of each fluorophore type were also used (GYACACACCGCCCG).

2.7. Polony microarray hybridizations

For each hybridization procedure, a pair of fluorescently labeled 10-mer probes was added to 6× SSPE at a final concentration of 10 nM, and 7 ml of this solution was added to each LockMailer Jar containing four polony microarrays. The jars were placed in an 87 °C water bath for 4 minutes to denature the single-stranded polony DNA, and then transferred to a RT water bath for 30 minutes for probe hybridization.

To wash the polony microarrays, the hybridization solutions were replaced with 3× SSC, and the jars are gently shaken for 15 or 45 minutes as follows: for most pairs, 15 minutes was sufficient, but pair 563-GAGGTGGAGC / 665-GCAAGTCGAA required a 45 minute wash. At the end of the wash time, the jars were immediately placed and then kept on ice until scanning is completed.

To scan the polony microarrays, a GenePix 4000b Microarray Scanner (Molecular Devices, Sunnyvale, CA) was used at 100% power at two wavelengths: 635 nm and 532 nm. The gain settings for each laser were adjusted for each probe-pair to achieve the maximum brightness possible without saturation of the colonies. All subsequent slides in an experiment were scanned at the same optimized gain setting. Each scan for each probe was saved as a 2969 × 4400-pixel image in 16-bit tiff format.

To remove the probes from the polony arrays, the 3× SSC solution was replaced with ultrapure water and the jars were placed in a 65 °C water bath for 7 minutes. The arrays

were then either hybridized with the next set of probes or stored overnight in Wash 1E (10 mM Tris (pH 7.5), 50 mM KCl, 2 mM EDTA, 0.01% Triton X-100) at 4 °C.

2.8. Data acquisition and processing

Raw intensities for each hybridization event were measured using ImageJ (National Center for Biotechnology Information, Bethesda, MD) and a plugin developed in our lab. All intensities were background subtracted. To adjust for polony size, the background-subtracted intensities were standardized by converting them into a percentage of the total for all probes.

Samples were analyzed using a community-level comparison method. UPGMA trees with jackknife support were created where each branch in a tree represents a whole bacterial community. Distances between bacterial communities were calculated using the Weighted UniFrac beta diversity metric (Lozupone et al. 2007) via QIIME (Caporaso et al. 2010). QIIME is a processing pipeline made for sequence data. A Euclidean distance matrix of the polony intensity vector averages from each cluster was created and a Newick-formatted tree file was created from it using the ctc package in R (Lucas and Jasson, 2006). To better visualize the 0-cycle and 35-cycle data from the 10-clone artificial bacterial community experiment, polony distance matrix values are also displayed in a principle component analysis using the cmdscale function in R (R Development Core Team, 2011).

3. Results and discussion

3.1. Polony OFRG analysis of an artificial bacterial community

To determine if a polony-based OFRG method could differentiate clones of various similarity from a known bacterial community, and to see if any affect from 35 cycles of PCR on the same community could be detected (versus 0 cycles of PCR), we constructed an artificial bacterial community comprised of ten different rRNA gene clones from eight genera and six phyla. These plasmids were used as the 0-cycle template, while the 35-cycle template was the 35-cycle PCR products of the 0-cycle template.

A polony OFRG analysis was performed on the 0- and 35-cycle templates from the 10-clone artificial bacterial community. A weighted UniFrac analysis from the resulting 11,800 polonies demonstrates a clear distinction between the two sample types versus the within sample replicates (Figure 2). The polonies are also depicted by a principal component analysis (Figure 3) in lieu of a dendrogram, which was too large for effective display. Here, polonies made from plasmid DNA (0-cycles of PCR) can be seen as ten relatively tight clusters of points (Figure 3A), demonstrating that all ten clones are distinguished with the 0-cycle template, even those at or near the species level (see Figure 1, clone-pairs 05/17 and 11/13).

Polonies made from the 35-cycle PCR products show a similar pattern of clustering (Figure 3B), yet the groups are much more diffuse than in Figure 3A, indicating changes to the DNA occurred during the 35 cycles of PCR. In addition, the clusters near the center are essentially merged and a new cluster appears to have formed in the lower-left quadrant (see arrow in Figure 3B).

There are at least two possible explanations for this more diffuse pattern. Both of them likely involve changes to the DNA relative to the original plasmid templates. PCR artifacts are known to occur in mixed-template reactions. One of those artifacts is the formation of chimeric sequences, where a partially duplicated amplicon from one species is fully extended from the template of one or more different species (Qiu et al., 2001; Wang and Wang, 1996). If this occurred during the PCR to create the 35-cycle templates, the hybridization patterns from them would also be chimeric, and therefore more variable than

the original 10 clones. Another potential source of DNA change that could explain the observed variation is point-mutations introduced by the polymerase itself during the 35 cycles of PCR. In this case, a polony arising from a mutated amplicon would have the same mutations (or more), potentially causing one or more probes to hybridize differently than the original DNA would have. Whatever artifacts may have occurred in the 35-cycle templates, the new cluster in Figure 3B (see arrow) suggests a specific type of artifact has occurred with some regularity under the conditions used.

3.2. Polony OFRG analysis of environmental bacterial communities

We also tested the ability of polony OFRG to differentiate bacterial communities from similar and different environments using a weighted UniFrac analysis (Figure 4). In this analysis, we analyzed human intestinal samples from subjects with Crohn's disease, ulcerative colitis and healthy individuals, as well as a sample of ocean water, which we expected to be considerably different than the human samples. One can see that the UniFrac distances between the different microbial communities are much greater than the distances between the three replicates within each community, indicating that not only can the communities be distinguished, but that consistency can be achieved with the method. For comparison, this analysis also included one of the 10-clone, 0-PCR- cycle replicates (topmost branch of Figure 4). As expected, the 10-clone sample exhibited the greatest divergence among the samples due to its simple makeup.

3.3. Conclusion and future directions

This proof-of-concept report demonstrates that polony OFRG, 1) can distinguish clones at or near the species level and higher, 2) can distinguish whole microbial communities in samples from similar or different environments, and 3) is repeatable, as shown by technical replicates.

The three key characteristics of polony microarrays that enable improvements to OFRG are, 1) the spatial isolation of DNA template before PCR, which reduces the formation of chimeras by reducing the intermingling of heterogeneous amplicons, 2) the ability to generate polony microarrays directly from genomic DNA, which transforms the confounding effect of PCR bias from one of population skew to one of skewed polony sizes, and 3) increased throughput, from 9,600-clone macroarrays (Bent et al., 2006) to microarrays potentially containing millions of polonies (Mitra and Church, 1999).

Future improvements could include the utilization of recently developed devices such the Polonator (Dover Systems, Salem) or Illumina sequencers (Illumina, Inc., San Diego, CA), which are flow-cell based systems with high-resolution fluorescence scanners. These two features would provide dramatic increases in throughput compared to our current slide-based method because they would enable analysis of higher density arrays as well as fully automated hybridizations and scans. Non-trivial protocol changes would be needed to adapt the machines to OFRG. For instance, if the two-dimensional bridge-amplification method currently used by Illumina is too inefficient for the longer amplicons used in OFRG, it may be necessary to switch to the 3-D hydrogel model used in this report. Additionally, to enable multiplexing of genomic DNA samples, flow cells would need to be partitioned at levels proportionate to the number of samples. Finally, should OFRG-friendly protocol changes be successfully implemented on these platforms and cluster formation from genomic DNA becomes possible, it would simultaneously confer the same benefits of reduced chimera formation and PCR skew to high-throughput sequencing of microbial communities.

Acknowledgments

The research is supported in part by NIH grant 5R01AI078885.

Abbreviations

OFRG	oligonucleotide fingerprinting of rRNA genes
QIIME	Quantitative Insights Into Microbial Ecology

References

- Barns SM, Fundyga RE, Jeffries MW, Pace NR. Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc. Natl. Acad. Sci. USA*. 1994; 91:1609–1613. [PubMed: 7510403]
- Bent E, Yin B, Figueroa A, Ye J, Fu Q, Liu Z, McDonald V, Jeske D, Jiang T, Borneman J. Development of a 9600-clone procedure for oligonucleotide fingerprinting of rRNA genes: utilization to identify soil bacterial rRNA genes that correlate in abundance with the development of avocado root rot. *J Microbiol Methods*. 2006; 67:171–180. [PubMed: 16712989]
- Borneman J, Chrobak M, Della Vedova G, Figueroa A, Jiang T. Probe selection algorithms with applications in the analysis of microbial communities. *Bioinformatics*. 2001; 17(Suppl 1):S39–48. [PubMed: 11472991]
- Borneman J, Triplett EW. Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. *Appl. Environ. Microbiol*. 1997; 63:2647–2653. [PubMed: 9212415]
- Butz JA, Yan H, Mikkilineni V, Edwards JS. Detection of allelic variations of human gene expression by polymerase colonies. *BMC Genet*. 2004; 5:3. [PubMed: 15040815]
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JL, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010; 7:335–336. [PubMed: 20383131]
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A*. 2011; 108(Suppl 1):4516–4522. [PubMed: 20534432]
- Chandler DP, Fredrickson JK, Brockman FJ. Effect of PCR template concentration on the composition and distribution of total community 16S rDNA clone libraries. *Mol. Ecol*. 1997; 6:475–482. [PubMed: 9161015]
- Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Heled J, Kearse M, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A. Geneious. 2010; v5.1 <http://www.geneious.com>.
- Edgar, Robert C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. 2004; 32(5):1792–97. [PubMed: 15034147]
- Farrelly V, Rainey FA, Stackebrandt E. Effect of genome size and rrn gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. *Appl. Environ. Microbiol*. 1995; 61:2798–2801. [PubMed: 7618894]
- Figueroa A, Borneman J, Jiang TJ. Clustering binary fingerprint vectors with missing values for DNA array data analysis. *J. Comput. Biol*. 2004; 11:887–901. [PubMed: 15700408]
- Koshkin AA, Singh SK, Nielsen P, Rajwanshi VK, Kumar R, Meldgaard M, Olsen CE, Wengel J. LNA (locked nucleic acids): synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation and unprecedented nucleic acid recognition. *Tetrahedron*. 1998; 54:3607–3630.
- Liu WT, Marsh TL, Cheng H, Forney LJ. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl. Environ. Microbiol*. 1997; 63:4516–4522. [PubMed: 9361437]

- Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol.* 2007; 73:1576–1585. [PubMed: 17220268]
- Lucas A, Jasson S. Using *amap* and *ctc* Packages for Huge Clustering. *R News.* 2006; 6:58–60.
- Mitra RD, Butty VL, Shendure J, Williams BR, Housman DE, Church GM. Digital genotyping and haplotyping with polymerase colonies. *Proc Natl Acad Sci U S A.* 2003; 100:5926–5931. [PubMed: 12730373]
- Mitra RD, Church GM. In situ localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Res.* 1999; 27:e34. [PubMed: 10572186]
- Muyzer G, De Waal ED, Uitterlinden AG. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction amplified genes coding for 16S rRNA. *Appl. Environ. Microbiol.* 1993; 59:695–700. [PubMed: 7683183]
- Obika S, Nanbu D, Hari Y, Andoh J, Morio K, Doi T, Imanishi T. Stability and structural features of the duplexes containing the nucleoside analogues with a fixed N-type conformation, 2'-O,4'-C-methylenerybonucleosides. *Tetrahedron Lett.* 1998; 39:5401–5404.
- Presley LL, Ye J, Li X, Leblanc J, Zhang Z, Ruegger PM, Allard J, McGovern D, Ippoliti A, Roth B, Cui X, Jeske DR, Elashoff D, Goodglick L, Braun J, Borneman J. Host-microbe relationships in inflammatory bowel disease detected by bacterial and metaproteomic analysis of the mucosal-luminal interface. *Inflamm Bowel Dis.* 2012; 18:409–417. [PubMed: 21698720]
- Price LB, Liu CM, Melendez JH, Frankel YM, Engelthaler D, Aziz M, Bowers J, Rattray R, Ravel J, Kingsley C, Keim PS, Lazarus GS, Zenilman JM. Community analysis of chronic wound bacteria using 16S rRNA gene-based pyrosequencing: impact of diabetes and antibiotics on chronic wound microbiota. *PLoS One.* 2009; 4:e6462. [PubMed: 19649281]
- Qiu X, Wu L, Huang H, McDonel P, Palumbo A, Tiedje J, Zhou J. Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rDNA gene-based cloning. *Appl. Environ. Microbiol.* 2001; 67:880–887. [PubMed: 11157258]
- R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2011. ISBN 3-900051-07-0, URL <http://www.R-project.org/>
- Reysenbach AL, Giver LJ, Wickham GS, Pace NR. Differential amplification of rRNA genes by polymerase chain reaction. *Appl. Environ. Microbiol.* 1992; 58:3417–3418. [PubMed: 1280061]
- Rieger C, Poppino R, Sheridan R, Moley K, Mitra R, Gottlieb D. Polony analysis of gene expression in ES cells and blastocysts. *Nucleic Acids Res.* 2007; 35:e151. [PubMed: 18073198]
- Ruegger PM, Della Vedova G, Jiang T, Borneman J. Improving probe set selection for microbial community analysis by leveraging taxonomic information of training sequences. *BMC Bioinformatics.* 2011; 12:394. [PubMed: 21985453]
- Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science.* 2005; 309:1728–1732. [PubMed: 16081699]
- Suzuki MT, Giovannoni SJ. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl. Environ. Microbiol.* 1996; 62:625–630. [PubMed: 8593063]
- Valinsky L, Della Vedova G, Scupham AJ, Alvey S, Figueroa A, Yin B, Hartin RJ, Chrobak M, Crowley DE, Jiang T, Borneman J. Analysis of bacterial community composition by oligonucleotide fingerprinting of rRNA genes. *Appl Environ Microbiol.* 2002; 68:3243–3250. [PubMed: 12089000]
- Wang GCY, Wang Y. The frequency of chimeric molecules as a consequence of PCR co-amplification of 16S rRNA genes from different bacterial species. *Microbiol.* 1996; 142:1107–1114.
- Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol.* 2007; 73:5261–5267. [PubMed: 17586664]
- Wu GD, Lewis JD, Hoffmann C, Chen YY, Knight R, Bittinger K, Hwang J, Chen J, Berkowsky R, Nessel L, Li H, Bushman FD. Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. *BMC Microbiol.* 2010; 10:206. [PubMed: 20673359]

- Yu H, Jeske DR, Ruegger P, Borneman J. Neutral Zone Classifiers Using a Decision-Theoretic Approach With Application to DNA Array Analyses. *J Agric Biol Environ Stat.* 2010; 15:474–490. [PubMed: 21769245]
- Zhang K, Zhu J, Shendure J, Porreca GJ, Aach JD, Mitra RD, Church GM. Long-range polony haplotyping of individual human chromosome molecules. *Nat Genet.* 2006; 38:382–387. [PubMed: 16493423]
- Zhu J, Shendure J, Mitra RD, Church GM. Single molecule profiling of alternative pre-mRNA splicing. *Science.* 2003; 301:836–838. [PubMed: 12907803]

Highlights

Polony OFRG distinguished 16S bacterial rRNA genes at or near the species level

Weighted UniFrac on polony OFRG data differentiates whole microbial communities

PCR artifacts are reduced via spatial isolation of genomic DNA prior to amplification

Figure 1.

Dendrogram of the near full-length 16S rRNA gene sequences used in the 10-clone artificial bacterial community with abbreviated taxonomic designations (CloneID.Phylum;Class;Genus) from the Ribosomal Database Project's Naïve Bayesian Classifier. The percentages after each taxonomic level indicate the confidence level of the classification. Branch lengths indicate the number of nucleotide changes per position.

Figure 2.

Weighed UniFrac analysis of OFRG-derived polonies from a 10-clone artificial bacterial community (A) 0-cycle and (B) 35-cycle templates. Jackknife support is indicated at the nodes. Polony counts for each of the three replicate microarrays were: 0-cycle, 1771, 2054, 1219; 35-cycle, 2050, 2287, 2419.

Figure 3.

Principal component analysis of OFRG-derived polonies from a 10-clone artificial bacterial community. Each spot represents a single polony, and the distance between each pair of polonies is a scaled Euclidean distance of their hybridization results: (A) 0-cycle and (B) 35-cycle templates. Polony counts for each of the three replicate microarrays were: 0-cycle, 1771, 2054, 1219; 35-cycle, 2050, 2287, 2419.

Figure 4.

Weighted UniFrac analysis of several environmental bacterial communities. Jackknife support is indicated at the nodes. Bacterial DNA used in the 13 microarrays that make up the trees came from five sources: ocean water, Crohn's disease (CD), ulcerative colitis (UC), healthy control (HC) and one artificial community of 10 bacterial phylotypes (10_clone). Microarray replicate numbers are indicated after label abbreviations. Polony counts for each of the three replicate microarrays were: Ocean, 1502, 1049, 1052; CD, 4051, 4685, 4457; UC, 1727, 1595, 1959; HC, 2428, 2239, 2174; 10_clone, 1771.