Journal of Microbiological Methods 105 (2014) 82–87

Contents lists available at ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

Improved resolution of bacteria by high throughput sequence analysis of the rRNA internal transcribed spacer

Paul M. Ruegger [a], Robin T. Clark [a], John R. Weger [b], Jonathan Braun [c], James Borneman [a,*]

[a] Department of Plant Pathology and Microbiology, University of California, Riverside, CA 92521, USA
[b] Department of Botany & Plant Sciences, University of California, Riverside, CA 92521, USA
[c] Department of Pathology and Lab Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA

ARTICLE INFO

Article history:
Received 18 June 2014
Accepted 2 July 2014
Available online 15 July 2014

Keywords:
High throughput sequencing
rRNA
Internal transcribed spacer
ITS
Illumina
Bacteria

ABSTRACT

Current high throughput sequencing (HTS) methods are limited in their ability to resolve bacteria at or below the genus level. While the impact of this limitation may be relatively minor in whole-community analyses, it constrains the use of HTS as a tool for identifying and examining individual bacteria of interest. The limited resolution is a consequence of both short read lengths and insufficient sequence variation within the commonly targeted variable regions of the small-subunit rRNA (SSU) gene. The goal of this work was to improve the resolving power of bacterial HTS. We developed an assay targeting the hypervariable rRNA internal transcribed spacer (ITS) region residing between the SSU and large-subunit (LSU) rRNA genes. Comparisons of the ITS region and two SSU regions using annotated bacterial genomes in GenBank showed much greater resolving power is possible with the ITS region. This report presents a new HTS method for analyzing bacterial composition with improved capabilities. The greater resolving power enabled by the ITS region arises from its high sequence variation across a wide range of bacterial taxa and an associated decrease in taxonomic heterogeneity within its OTUs. Although the method should be adaptable to any HTS platform, this report presents PCR primers, amplification parameters, and protocols for Illumina-based analyses.

1. Introduction

New and important experimental capabilities for examining bacterial communities have come from application of high throughput sequencing (HTS) technology to the small-subunit rRNA (SSU) gene. The combination of large numbers of sequencing reads and multiplexing technologies provides a cost-effective method to perform in-depth analysis of large numbers of samples. These capabilities facilitate investigations examining temporal and spatial variables (Caporaso et al., 2011a; Costello et al., 2009), and are fostering a broader understanding of the roles that bacterial communities play in ecosystem functioning.

Although current HTS methods are limited in their ability to taxonomically classify bacterial sequences at or below genus levels (Caporaso et al., 2011b; Price et al., 2009a,b; Wu et al., 2010), a related but less considered weakness is their limited ability to resolve such bacterial sequences. In this report, resolution, or resolving power, is defined as the ability to distinguish between different types of bacteria by their sequence differences, irrespective of whether taxonomy has been or can be imputed to them from the outset. Greater resolving power enables various analyses to be performed at finer levels, enhancing the ability to, for example, correlate specific bacteria with host or environmental traits, monitor population shifts, or assess microgeographical relationships through the use of sequence-selective PCR primers and probes.

To achieve our goal of improving the resolving power of bacterial HTS, we developed an assay that targets the rRNA internal transcribed spacer (ITS) region. Because of its high sequence variation, the ITS region has been used for species and subspecies-level classification of bacteria (Barry et al., 1991; Frothingham and Wilson, 1993; McLaughlin et al., 1993) and for ribosomal internal spacer analyses (RISA/ARISA) of community composition (Borneman and Triplett, 1997; Fisher and Triplett, 1999). Experimental protocols were created to facilitate the use of our new ITS HTS method, including PCR primers, amplification parameters and protocols for Illumina-based analyses. Using annotated bacterial genomes, the resolving power of our new method was assessed by a comparative analysis of intra-region sequence differences in the ITS region and two commonly used SSU regions. While our method's primary strengths do not rely on taxonomically identifying sequences it can be useful to do so. Therefore, we created an SSU-ITS sequence database and compare the accuracy of taxonomic assignments between the ITS

Abbreviations: ARISA, automated ribosomal intergenic spacer analysis; HTS, high throughput sequencing; ITS, internal transcribed spacer; LSU, large subunit; OTU, operational taxonomic unit; PCoA, principle coordinates analysis; RISA, ribosomal intergenic spacer analysis; SSU, small subunit; SSU-ITSdb, a sequence database was created from annotated bacterial genomes to contain both SSU and ITS regions; SSU-ITSpt, a phylogenetic tree used to impute phylogeny to ITS and SSU reads for performing comparative UniFrac analyses on both regions.

* Corresponding author. Tel.: +1 951 827 3584.
E-mail address: borneman@ucr.edu (J. Borneman).

http://dx.doi.org/10.1016/j.mimet.2014.07.001
0167-7012/

and SSU regions in a mock community analysis. Moreover, we demonstrate a feasible way of performing a phylogenetic beta diversity analysis using ITS reads in a proof-of-concept experiment. After constructing a phylogenetic tree from the SSU genes of the SSU-ITS sequence database we used it to compare the experimental results obtained from ITS and SSU assays of the fecal bacteria of two mouse genotypes.

2. Materials and methods

2.1. Sample acquisition and DNA extraction

Fecal samples were collected from mice of two genotypes, obtained from Jackson Laboratories (Bar Harbor, Maine, USA): C3H-IL10-KO = C3Bir.129P2 (B6)-Il10tm1Cgn/Lt and C3H = C3H/HeJ. Animal use protocols were approved by the Institutional Animal Care and Use Committee at the University of California, Riverside. DNA was extracted from the samples using the PowerSoil DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA, USA), and a 30-second beat-beating step using a Mini-Beadbeater-16 (BioSpec Products, Bartlesville, OK, USA). High throughput sequencing analysis of bacterial rRNA genes was performed on the resulting DNA using two Illumina-based assays: one targeting the small-subunit (SSU) rRNA gene and the other targeting the rRNA internal transcribed spacer (ITS) region.

2.2. SSU assay

One hundred microliter amplification reactions were performed in an MJ Research PTC-200 thermal cycler (Bio-Rad Inc, Hercules, CA, USA) containing: 50 mM Tris (pH 8.3), 500 μg/ml bovine serum albumin (BSA), 2.5 mM MgCl2, 250 μM of each deoxynucleotide triphosphate (dNTP), 400 nM of each PCR primer, 4 μl of DNA template, and 2.5 units JumpStart Taq DNA polymerase (Sigma-Aldrich, St. Louis, MO, USA). The PCR primers targeted a portion of the SSU rRNA gene containing the hypervariable V4 region, with the reverse primers including a 12-bp barcode (Table S1) (Caporaso et al., 2011b). We note that the 3′ end of the forward PCR primer used in this assay (515F, GTGCCAGCMGCCGCGGTAA) is identical to the one used in the bioinformatics analyses described in this report (SSU-517F, GCCAGCAGCCGCGGTAA); thus, amplification for both primers begins at the same place. PCR primers were only frozen and thawed once. Thermal cycling parameters were 94 °C for 5 min; 35 cycles of 94 °C for 20 s, 50 °C for 20 s, and 72 °C for 30 s, and followed by 72 °C for 5 min. PCR products were purified using the MinElute 96 UF PCR Purification Kit (Qiagen, Valencia, CA, USA). DNA sequencing was performed using an Illumina HiSeq 2000 (Illumina, Inc, San Diego, CA). Clusters were created using template concentrations of 1.9 pM and PhiX at 1.37 pM; PhiX is recommended by the manufacturer for samples with uneven distributions of A, C, G and T. One hundred base sequencing reads of the 5′ end of the amplicons and seven base barcode reads were obtained using the sequencing primers listed in Table S1.

2.3. ITS assay

One hundred microliter amplification reactions were performed in an MJ Research PTC-200 thermal cycler (Bio-Rad Inc, Hercules, CA, USA) containing: 50 mM Tris (pH 8.3), 500 μg/ml bovine serum albumin (BSA), 2.5 mM MgCl2, 250 μM of each deoxynucleotide triphosphate (dNTP), 400 nM of the forward PCR primer, 200 nM of each reverse PCR primer, 4 μl of DNA template, and 2.5 units JumpStart Taq DNA polymerase (Sigma-Aldrich, St. Louis, MO, USA). PCR primers targeted a portion of the SSU and LSU rRNA genes and the hypervariable internal spacer region, with the reverse primers including a 12-bp barcode (Table S2); primer binding sites are the reverse and complement of the commonly used SSU primer 1492R (Frank et al., 2008) and LSU primer 129F (Hunt et al., 2006). PCR primers were only frozen and thawed once. Thermal cycling parameters were 94 °C for 5 min; 35 cycles of 94 °C for 20 s, 56 °C for 20 s, and 72 °C for 40 s, and followed by 72 °C for 5 min. PCR products were purified using the MinElute 96 UF PCR Purification Kit (Qiagen). DNA sequencing was performed using an Illumina HiSeq 2000 (Illumina, Inc). Clusters were created using template concentrations of 2.5 pM. Although we did not add PhiX before cluster formation in this assay, its use may improve sequence quality. One hundred base sequencing reads of the 5′ end of the amplicons and seven base barcode reads were obtained using the sequencing primers listed in Table S2.

2.4. Sequence processing

QIIME (Caporaso et al., 2010b) was used to process reads from both ITS and SSU Illumina assays, which contained a total of 85 and 80 samples, respectively, from various microbial habitats in mice. For this study, we chose the largest subsets of comparable samples between the two assays, which were fecal samples originating from the same 14 mice (seven each of C3H and C3H-IL10-KO). De-multiplexing and removal of low-quality sequences was performed with default parameters of: minimum Q-score = 3, maximum number of N characters allowed = 0, maximum number of consecutive low-quality base calls allowed before truncating a read = 3. The minimum number of consecutive high-quality base calls to include a read as a fraction of input length = 0.99, which is higher than the default of 0.75. All filtered reads had a length of 101 bp. For the ITS assay, the total initial number of sequencing reads was 71,281,079. Numbers of sequences removed using the aforementioned quality control parameters were: barcode errors (3,923,456), reads too short after quality truncation (6,728,482), and too many Ns (21,520), leaving a total of 60,607,621 filtered reads. For the 14 fecal samples used in this study, filtered read counts ranged from 312,405 to 1,403,065. For the SSU assay, the total initial number of sequencing reads was 66,148,010. Numbers of sequences removed using the aforementioned quality control parameters were: barcode errors (3,559,863), reads too short after quality truncation (7,118,807), and too many Ns (16,866,051), leaving 38,603,289 filtered reads. For the 14 fecal samples used in this study, filtered read counts ranged from 384,301 to 603,227.

2.5. Simulated PCR

Simulated PCR was performed on annotated bacterial genome sequences with Geneious 6.1 (Biomatters Ltd) using the primers and number of allowed mismatches listed in Table 1.

2.6. Resolving power

For each region and simulated PCR amplicon length (100 bp and 400 bp), we calculated the average pairwise sequence difference of amplicons arising from all species within each genus, and all subspecies within each species. Only bacteria having at least two species were selected. Differences were counted at equal weighting after pairwise alignment, and included any character difference between aligned sequences. Resolving power for OTUs was determined using the original taxonomy of the representative sequence from each OTU.

Table 1
Primers for simulated PCR.

Primer name	Sequence	Length	Allowed mismatches
SSU-341F	CCTACGGGNGGCNGCA	16	2
SSU-517F	GCCAGCAGCCGCGGTAA	17	2
SSU-1406R	GACGGGCGGTGWGTRCA	17	2
ITS-1507F	GGTGAAGTCGTAACAAGGTA	20	3
ITS-23SR	GGGTTBCCCCATTCRG	16	2

SSU-1406R was used as the reverse primer for both SSU forward primers. The criteria for a successful hybridization of a primer to genomic DNA was: mismatches ≤ allowed mismatches, length of match = primer length.