Journal of Applied & Environmental Microbiology, 2023, Vol. 11, No. 1, 1-10
Available online at http://pubs.sciepub.com/jaem/11/1/1
Published by Science and Education Publishing
DOI:10.12691/jaem-11-1-1

Independent Validation of Differential Abundance Patterns from Illumina Miseq Analysis Using Quantitative PCR Techniques on the Selective Primer for *Chitinophaga*

Spencer Debenport[1], Laura Mason[2], Richard P Dick[2,*]

[1]Department of Plant Pathology, The Ohio State University-OARDC, Wooster, OH
[2]School of Environment and Natural Resources, The Ohio State University, Columbus, OH, USA
*Corresponding author: dick.78@osu.edu

Received March 12, 2023; Revised April 17, 2023; Accepted April 26, 2023

Abstract A criticism of amplicon sequencing is the potential for bias during PCR amplification. Quantitative PCR (qPCR) is an independent validation that can estimate taxon abundance and confirm patterns observed in amplicon sequencing patterns. Therefore, the objective was to design primers based on NGS sequencing and test qPCR primers to validate abundance patterns of bacterial and fungal OTUs on soils from the Optimized Shrub-intercropping System (OSS) or Sole Cropping in the Sahel. The results showed that quantitative PCR (qPCR) independently validated patterns observed in high throughput sequencing (HTS) analyses. Specific sub-genus level OTU clusters were found to be significantly enriched in intercropped millet plants in an experiment using the Ilumina MiSeq platform. These OTU sequences were used to design primers to independently validate the trends observed in that study. A total of seven OTU clusters were targeted in the Aspergillus, Chitinophaga, Fusarium, Lasiodiplodia, and Penicillium genera. The majority of those primers showed poor specificity for their intended targets, while the Chitinophaga specific primer set showed clear amplification with a single band at the expected size. This primer was used for qPCR analysis of the same DNA templates used for the Illumina MiSeq study. Quantitative PCR shows significant ($P < 0.05$) enrichment of Chitinophaga marker DNA that match the previously observed patterns. MiSeq analysis showed two times higher fold change differences in markers than observed in the qPCR study. These results demonstrate that selective primers can be designed from OTU sequence data and that qPCR analysis can be utilized to independently validate trends observed in HTS studies.

Keywords: *Optimized Shrub-intercropping System (OSS), Operational Taxonomic Unit (OTU), Quantitative Polymerase Chain Reaction (qPCR), High-throughput Sequencing (HTS), Plant Growth Promoting Rhizobacteria (PGPR)*

Cite This Article: Spencer Debenport, Laura Mason, and Richard P Dick, "Independent Validation of Differential Abundance Patterns from llumina Miseq Analysis Using Quantitative PCR Techniques on the Selective Primer for *Chitinophaga*." *Journal of Applied & Environmental Microbiology*, vol. 11, no. 1 (2023): 1-10. doi: 10.12691/jaem-11-1-1.

1. Introduction

A reduction of costs and improved ease of analysis has made high throughput sequencing techniques available to microbial ecologists working in a variety of systems and scales. The large sample numbers that can be processed together allows researchers to analyze factors controlling soil microbial community composition at the global scale without culturing [1]. Broad conclusions can be drawn from a multitude of studies, whether it be the influence of a plant species on shaping its rhizosphere microbial community [3] or the effect of soil inoculation with a biocontrol organism on the total microbial community [3].

However, a key criticism of amplicon sequencing is the potential biases introduced during initial PCR amplification step [4], and this critique must be addressed through direct validation of changes in microbial biomass via quantitative cell or DNA-based methods [5]. Quantitative PCR (qPCR) has been suggested as a method for performing this independent validation, and as a means of providing an accurate estimation of taxon abundance and confirm patterns observed in amplicon sequencing results [5,6].

The design and optimization of sequence-specific primers within marker gene regions is difficult; rigorous testing must also be performed to ensure optimal PCR conditions and primer binding specificity [7,8,9]. Here, in-depth sequence knowledge of the region to be amplified

is helpful in designing taxon-specific primers. For example, Kwak et al [6] used amplicon sequencing data from the 16S rRNA V1 – V3 region to discover a strain of Flavobacter, and using this knowledge, designed qPCR primers for this specific isolate to determine copy numbers in the rhizosphere of wilt-susceptible and wilt-resistant tomato cultivars, confirming results from their NGS experiments. Yang et al [10], using similar validation methods in a soil microbiome study, identified patterns of enrichment for certain microbial taxa associated with peach replant disease.

Our study had a comparable objective to design and test qPCR primers and then validate differential abundance patterns of specific OTU clusters that were observed by Debenport et al. [11]. That study had differences in abundance for certain bacterial and fungal OTUs between Optimized Shrub Intercropping systems (+OSS) and sole cropping systems (-OSS) in the Sahel. The specific objective was to design these primers based on NGS sequencing results from the differentially abundant amplicons discovered by Debenport et al [11]. The current study takes the next logical step by quantitatively verifying results.

2. Materials and Methods

2.1. Soil Samples

Soil samples came from two long term study sites in the prime cropping region of Senegal, the Peanut Basin, near Keur Matar Senegal (14°45 N, 16°51 W) and Nioro (13°45 N, 15°47 W) with annual rainfall regimes of 450 and 750 mm, and soil types of Dior soil (Rubic arenosol with 95% sand) and Deck-Dior soil (Haplic Ferric Lixisol with around 90% sand), respectively. In brief, both experiments had a Randomized Complete Block Split Plot Design with the Optimized Shrub-intercropping System (OSS) as the main plot (+ and -OSS) and NPK fertilizer rate (0 to 1.5 recommended NPK rate) as the sub-plot plot treatments. Keur Matar was intercropped with *Guiera senegalensis*, whereas Nioro was intercropped with *Pilio reticulatum*. The experiment was established in 2003 with OSS treatment having 1400-1833 shrubs ha^{-1} and annual incorporation of coppiced aboveground biomass. Otherwise, the experiments followed local farmer practices with the common rotation of peanut (*Arachis hypogaea* var 55-437) and millet (*Pennissetum glaucum* var Souna 3). Further details of these experiments can be found at Debenport et al., (2015) [11] and Bright et al. (2017, 2021) [12,13].

On August 13, 2013, and September 10, 2013 at the Nioro and Keur Matar sites, respectively, the zero-fertilizer rate sub-plots of + and – OSS were sampled. This treatment follows the common practice of Senegalese farmers who rarely use chemical fertilizers. Samples from two independent transects (Sampling 1 and 2) were taken at each location where each sampling had 4 field replications. Soil samples taken with a 2.5 cm diameter probe (0 to 20 cm depth) from root zones of 5 millet plants (grown either in Optimized Shrub-intercropping or Sole Cropping) were composited by homogenization and passing a 2 mm sieve.

2.2. DNA Extraction and Illumina MiSeq Sequencing

DNA was extracted from 0.25 g of each soil sample using MoBio PowerSoil DNA Isolation Kits (MoBio Laboratories, Carlsbad, CA). 50µl of extracted DNA was precipitated to pellet form using 3M sodium acetate for transport to Wooster, OH. Upon arrival, DNA pellets were resuspended in 50µl Qiagen (Qiagen, Venlo, Netherlands) elution buffer and stored in a -20°C freezer. Amplicon libraries for the 16S rRNA V4 region were generated using methods outlined by 14. Amplicon libraries for the ITS1 region were generated using the protocol described in 15. For each of these library preparation methods, we created a single PCR replicate instead of triplicate replicates in the listed protocols. For each amplicon type, all sample libraries were pooled together and purified using a Pippen Prep instrument (Sage Science, Beverly, MA). Purified library pools were quantified using the Qubit double-stranded DNA high-sensitivity assay (Life Technologies, Guilford, CT). Amplicon libraries were sequenced on an Illumina MiSeq instrument using 250 base paired-end kits at the Molecular and Cellular Imaging Center at the Ohio State University Ohio Agricultural Research and Development Center in Wooster, OH.

16S rRNA amplicon sequence processing. The 16S sequence paired-end data set was demultiplexed on the MiSeq instrument itself at the time of sequencing. Each pair of reads was joined and quality filtered using the 'join_paired_ends.py' script provided in the Quantitative Insights into Microbial Ecology (QIIME1) software suite [14] (Caporaso et al., 2010). We utilized the open reference operational taxonomic unit (OTU) picking protocol in QIIME1. Briefly, sequences were clustered against the 2013 Greengenes ribosomal database 97% reference dataset (http://greengenes.secondgenome.com/downloads). Sequences which did not match any entries in this reference were subsequently clustered into de novo OTUs at 97% similarity with UCLUST. Taxonomy was assigned to all OTUs using the RDP classifier (Michigan State University, http://rdp.cme.msu.edu/classifier/classifier.jsp) within QIIME1, again using the Greengenes reference dataset.

ITS1 amplicon sequence processing. The ITS data set was demultiplexed on the MiSeq instrument itself at the time of sequencing. Not all of the paired end reads were overlapping due to the variable length of the ITS1 region, so we used the forward read from each sequence for our downstream analysis. Reads were clustered into OTUs using the same open reference OTU picking protocol in QIIME1 as the 16S data set, using the UNITE+INSD (International Nucleotide Sequence Databases; NCBI, EMBL, DDBJ) 97% reference database. Taxonomy was assigned to all OTUs using the RDP classifier (Michigan State University, http://rdp.cme.msu.edu/classifier/classifier.jsp) within QIIME1, again using the UNITE+INSD reference dataset.

Determining Differentially abundant OTUs. Agronomic, soil, and sequence count data were subjected to analyses of variance using the general linear model and Tukey's test in R [15] (R core Team, 2014). A two-factor model consisting of treatment and block, with four replicated blocks at each site was used. At each site, two experiments

with a randomized complete block design and four blocks were conducted. Comparisons of individual OTU abundances were performed across soil treatments (+OSS vs. -OSS) for each transect at each site. OTU tables containing read counts for each OTU in each sample, taxonomy information for each OTU, and sample metadata for each sample were exported from QIIME and imported into R using the Phyloseq R package [16] (McMurdie and Holmes, 2013). OTUs representing < 0.001% of the total number of sequences from each library were removed. OTU tables were subset for each comparison and formatted for the DESeq2 [17] (Love et al., 2014) package in R.

Differential abundance of OTUs by sample type was determined using the DESeq2 R package. Sequences for all OTUs belonging to genera shared commonly between experiments were extracted from representative sequence files in QIIME. Sequences for each genus were aligned using ClustalW [18] (Thompson et al., 1994) and trees were constructed using the neighbor-joining method with 1,000 bootstrap replicates in Geneious 6.0.3 (Biomatters, available from http://www.geneious.com). Groupings of OTUs within a genus were classified by bootstrap values over 90. These groupings were used to design primers for library validation via qPCR ($P < 0.05$).

2.3. OTU Cluster Specific qPCR Primer Design.

Primers for quantitative PCR (qPCR) were designed from 16S and ITS amplicon libraries described above, and specific targets are thoroughly described in [11] Debenport et al. (2015). Briefly, clusters of OTUs found to be consistently enriched (30-fold) in intercropped millet samples included members of the following: *Chitinophaga*, *Aspergillus*, *Coniella*, *Epicoccum*, *Fusarium*, *Gibberella*, *Lasiodiplodia*, *Penicillum*, and *Phoma*. Chitinophaga [11] (Debenport et al 2015). OTU sequence lengths of around 250 bases were used to identify variable regions specific to clusters. Variable regions were identified as specific to OTU clusters while not matching to the rest of identified sequences within those genera. OTU sequences were aligned using Geneious 6.0.3 software (Biomatters, available from http://www.geneious.com) and variable regions unique to targeted sub-genus clusters were identified. Primers are designed to encompass these variable regions with minimal matching to the remaining OTUs within each genus. A *Chitinophaga* specific primer was matched with the universal 16S V4 region primer described in Caporaso et al. (2012) [19]. ITS OTU cluster specific primers were matched with universal ITS1-F and ITS2 primers described in [20] Gardes and Bruns (1993), which were used to generate the OTU sequences in [11] Debenport et al., 2015. Specific primers were designed to allow for ~100-200 base amplicons when paired with universal primers.

Initial testing of qPCR primer sets. Initial quantity of DNA template was determined by Qubit double-stranded DNA high-sensitivity assay (Life Technologies, Guilford, CT) [11] (Debenport et al., 2015). Polymerase chain reactions (PCRs) were performed with each primer set using 2 µl template DNA, 0.2 µl 100 mM primers, 5 µl 5x GoTaq Buffer, 1.8 µl 25 mM MgCl2, 2.5 µl 2mM dNTPs, 13 µl sterile water, and 0.3 µl GoTaq polymerase (Promega, Madison, WI). PCR was run at 94°C for 3 minutes, 94°C for 45 seconds, 53°C for one minute, 72°C for 90 seconds for 30 cycles. A 10 minute extension at 72°C was used after this. Confirmation of successful PCR was performed through gel electrophoresis using a 5 µl PCR product run on a 2% agarose gel at 120V for 1 hour. Band intensity was determined using densiometric scans of gel images using ImageJ software (W Rasband, NIH, Bethesda, MD).

2.4. Quantitative PCR Analysis of *Chitinophaga* 16S Marker

Quantitative PCR was performed using the Chit and 16Sr primer set. For each reaction 12.5 µl SYBR GreenMaster Mix (Life Technologies, Grand Island, NY), 10.3 µl water, 0.1 µl 100 mM of each primer, and 2 µl of template DNA at a 1:30 dilution. Reactions were run at 95°C for 3 minutes, 95°C for 10 seconds, 53°C for 30 seconds, with 40 total cycles. PCR product from a reaction using the Chit and 16Sr primer set as described above was used to create a standard for the qPCR. This product was purified using a Wizard PCR Purification kit (Promega, Madison, WI). A ten-fold serial dilution of this product was used for the standard curve. Initial purified PCR product was measured using a Nanodrop spectrophotometer (Thermo Scientific, Wilmington, DE) and a standard curve with DNA concentrations at 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, and 0 ng/µl was used. Initial quantity of target DNA in template DNA was estimated using this standard curve.

2.5. Statistical Analyses

Threshold cycle number and log transformed starting DNA quantifications were checked for statistical significance using the general linear model in R [15] (R Core Team 2014). Results were marked as significant at the $P < 0.05$ level.

3. Results

3.1. OTU Cluster Specific qPCR Primer Design

A total of seven OTU clusters previously identified were selected for specific primer design [11] (Debenport et al., 2015). These clusters were spread across the *Aspergillus*, *Chitinophaga*, *Fusarium*, *Lasiodiplodia*, and *Penicillium* genera with *Aspergillus* and *Penicillium* containing two clusters each. Variable region location and composition in these clusters led to limited options for cluster specific primer design. Primer locations and directionalities are shown in Figure 1. The reverse compliment sequence for Pen1 was paired with ITS1-F primer to allow for a longer amplicon size.

Primer targets and sequences are shown in Table 1.

3.2. Validation of Primer Specificity

All seven OTU cluster specific primer sets were tested to confirm amplification of target sequences. The majority of primer sets showed smearing or multiple banding patterns on gel images when run with the standard PCR protocol (Figure 2). The smearing and multiple banding pattern indicates that these primers are bound to non-target DNA or that the PCR conditions are not conducive for amplification of target DNA. Of all tested primers, the Chit primer resulted in one specific amplicon of the correct sequence length (Figure 3), and brighter band intensity (130% higher) in intercropped millet samples compared to bare soil millet samples. Due to the presence of one clear band at the estimated size, the Chit primer set was chosen to move forward to qPCR analysis.

3.3. Quantitative PCR Validation

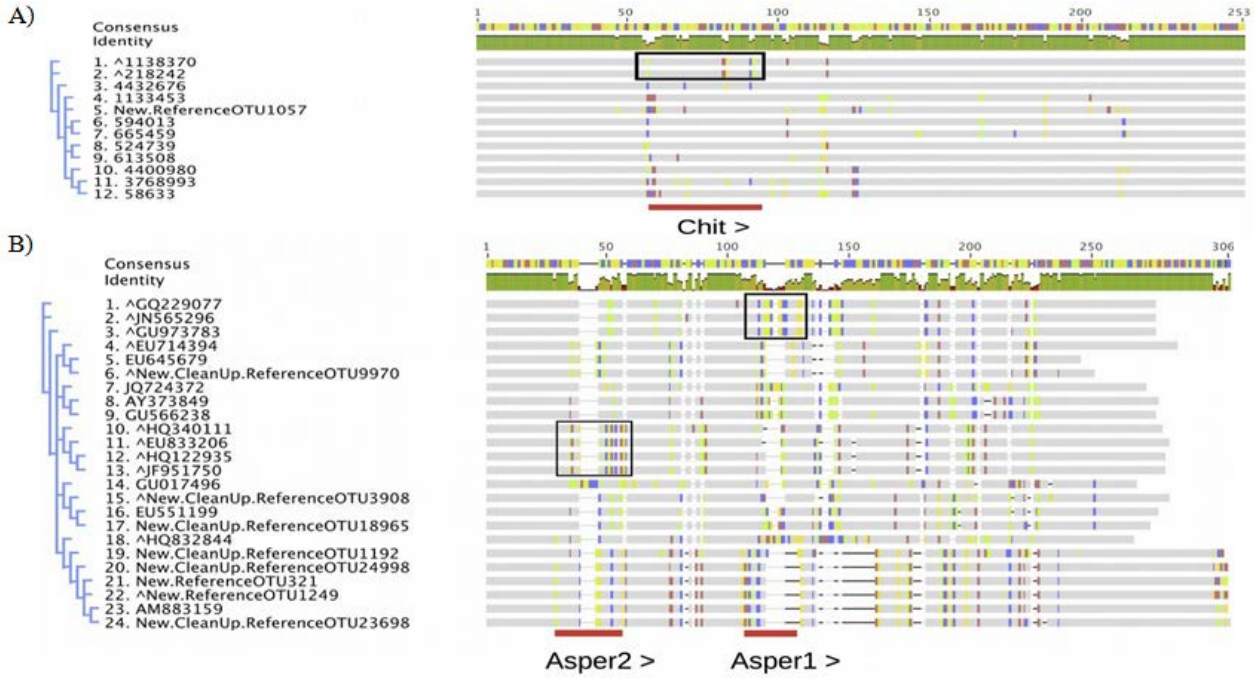
Both the threshold cycle number and starting DNA quantity for *Chitinophaga* markers were significantly ($P < 0.05$) higher in intercropped millet samples than in millet grown in bare soil samples (Figure 4), validating previous NGS results [11] (Debenport et al., 2015). Fold change differences in targeted *Chitinophaga* OTUs were calculated from the estimated starting DNA quantity (Table 2). These fold changes were compared to those calculated in the MiSeq analysis [11] (Debenport et al., 2015) using these same DNA templates. Log 2-fold change values were on average two twice as large in the MiSeq analysis compared to those calculated using qPCR.

Table 1. Description of OT U cluster-specific primers used in qPCR assays

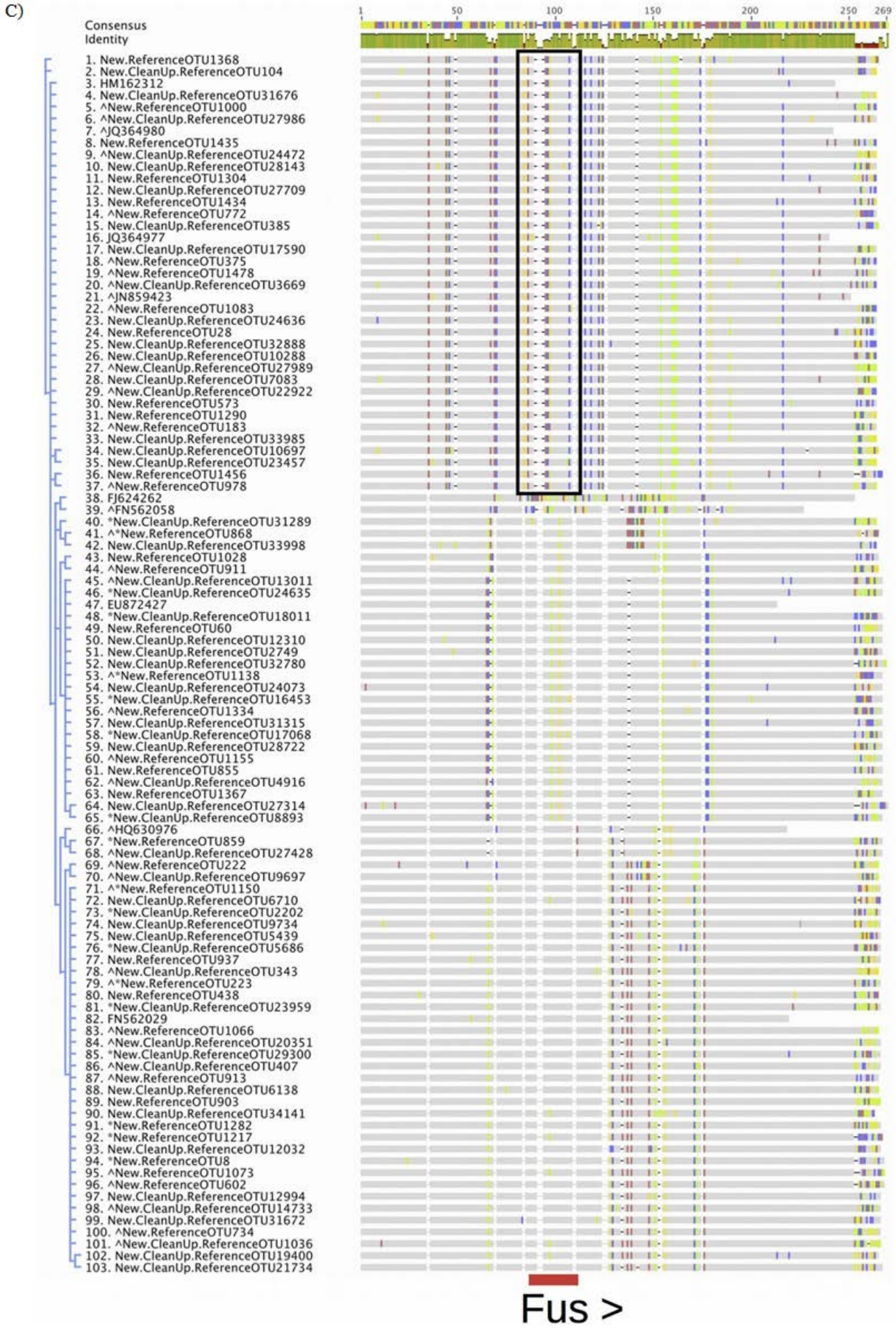
Primer Name	Primer Sequence	Tm (ºC)	Target†
Asper1	CCG CTT GTC GGC CGC CGG G	70	*Aspergillus* cluster A
Asper2	GAG TGT AGG GTT CCT AGC GAG C	59	*Aspergillus* cluster D
Chit	GTG AAA TCT CCA GGC TTA ACC T	53	*Chitinophaga* cluster B
Fus	GAA CAG ACG GCC CCG TAA CAG G	60	*Fusarium* cluster A
Lasio	ACA CCT CTG TTG CCT CGG C	62	*Lasiodiplodia* cluster A
Pen1	AGC CCA TCT TCA GGG TTC A	55	*Pennicillium* cluster A
Pen2	TAT CGT ACC TTG CTT CGG C	55	*Pennicillium* cluster B
806r	GGA CTA CHV GGG TWT CTA AT	45	Universal 16S V4 region
ITS2	GCT GCG TTC TTC ATC GAT GC	54	Universal ITS2 region§
ITS1-F	CTT GGT CAT TTA GAG GAA GTA A	49	Universal ITS2 region§

†Subgenus level clusters are described in Debenport et al. (2015).

§ Described in White et al. (1990).



C)


Consensus
Identity
1. New.ReferenceOTU1368
2. New.CleanUp.ReferenceOTU104
3. HM162312
4. New.CleanUp.ReferenceOTU31676
5. ^New.ReferenceOTU1000
6. ^New.CleanUp.ReferenceOTU27986
7. ^JQ364980
8. New.ReferenceOTU1435
9. ^New.CleanUp.ReferenceOTU24472
10. New.CleanUp.ReferenceOTU28143
11. New.ReferenceOTU1304
12. New.CleanUp.ReferenceOTU27709
13. New.ReferenceOTU1434
14. ^New.ReferenceOTU772
15. New.CleanUp.ReferenceOTU385
16. JQ364977
17. New.CleanUp.ReferenceOTU17590
18. ^New.ReferenceOTU375
19. ^New.ReferenceOTU1478
20. ^New.CleanUp.ReferenceOTU3669
21. ^JN859423
22. ^New.ReferenceOTU1083
23. New.CleanUp.ReferenceOTU24636
24. New.ReferenceOTU28
25. New.CleanUp.ReferenceOTU32888
26. New.CleanUp.ReferenceOTU10288
27. ^New.CleanUp.ReferenceOTU27989
28. New.CleanUp.ReferenceOTU7083
29. ^New.CleanUp.ReferenceOTU22922
30. New.ReferenceOTU573
31. New.ReferenceOTU1290
32. ^New.ReferenceOTU183
33. New.CleanUp.ReferenceOTU33985
34. New.CleanUp.ReferenceOTU10697
35. New.CleanUp.ReferenceOTU23457
36. New.ReferenceOTU1456
37. ^New.ReferenceOTU978
38. FJ624262
39. ^FN562058
40. *New.CleanUp.ReferenceOTU31289
41. ^*New.ReferenceOTU868
42. New.CleanUp.ReferenceOTU33998
43. New.ReferenceOTU1028
44. ^New.ReferenceOTU911
45. ^New.CleanUp.ReferenceOTU13011
46. *New.CleanUp.ReferenceOTU24635
47. EU872427
48. *New.CleanUp.ReferenceOTU18011
49. New.ReferenceOTU60
50. New.CleanUp.ReferenceOTU12310
51. New.CleanUp.ReferenceOTU2749
52. New.CleanUp.ReferenceOTU32780
53. ^*New.ReferenceOTU1138
54. New.CleanUp.ReferenceOTU24073
55. *New.CleanUp.ReferenceOTU16453
56. ^New.ReferenceOTU1334
57. New.CleanUp.ReferenceOTU31315
58. *New.CleanUp.ReferenceOTU17068
59. New.CleanUp.ReferenceOTU28722
60. ^New.ReferenceOTU1155
61. New.ReferenceOTU855
62. ^New.CleanUp.ReferenceOTU4916
63. New.ReferenceOTU1367
64. New.CleanUp.ReferenceOTU27314
65. *New.CleanUp.ReferenceOTU8893
66. ^HQ630976
67. *New.ReferenceOTU859
68. ^New.CleanUp.ReferenceOTU27428
69. ^New.ReferenceOTU222
70. ^New.CleanUp.ReferenceOTU9697
71. ^*New.ReferenceOTU1150
72. New.CleanUp.ReferenceOTU6710
73. *New.CleanUp.ReferenceOTU2202
74. New.CleanUp.ReferenceOTU9734
75. New.CleanUp.ReferenceOTU5439
76. *New.CleanUp.ReferenceOTU5686
77. New.ReferenceOTU937
78. ^New.CleanUp.ReferenceOTU343
79. ^*New.ReferenceOTU223
80. New.ReferenceOTU438
81. *New.CleanUp.ReferenceOTU23959
82. FN562029
83. ^New.ReferenceOTU1066
84. ^New.CleanUp.ReferenceOTU20351
85. *New.CleanUp.ReferenceOTU29300
86. ^New.CleanUp.ReferenceOTU407
87. ^New.ReferenceOTU913
88. New.CleanUp.ReferenceOTU6138
89. New.ReferenceOTU903
90. New.CleanUp.ReferenceOTU34141
91. *New.ReferenceOTU1282
92. *New.ReferenceOTU1217
93. New.CleanUp.ReferenceOTU12032
94. *New.ReferenceOTU8
95. ^New.ReferenceOTU1073
96. ^New.ReferenceOTU602
97. New.CleanUp.ReferenceOTU12994
98. ^New.CleanUp.ReferenceOTU14733
99. New.CleanUp.ReferenceOTU31672
100. ^New.ReferenceOTU734
101. ^New.CleanUp.ReferenceOTU1036
102. New.CleanUp.ReferenceOTU19400
103. New.CleanUp.ReferenceOTU21734
1
50
100
150
200
250
269
Fus >

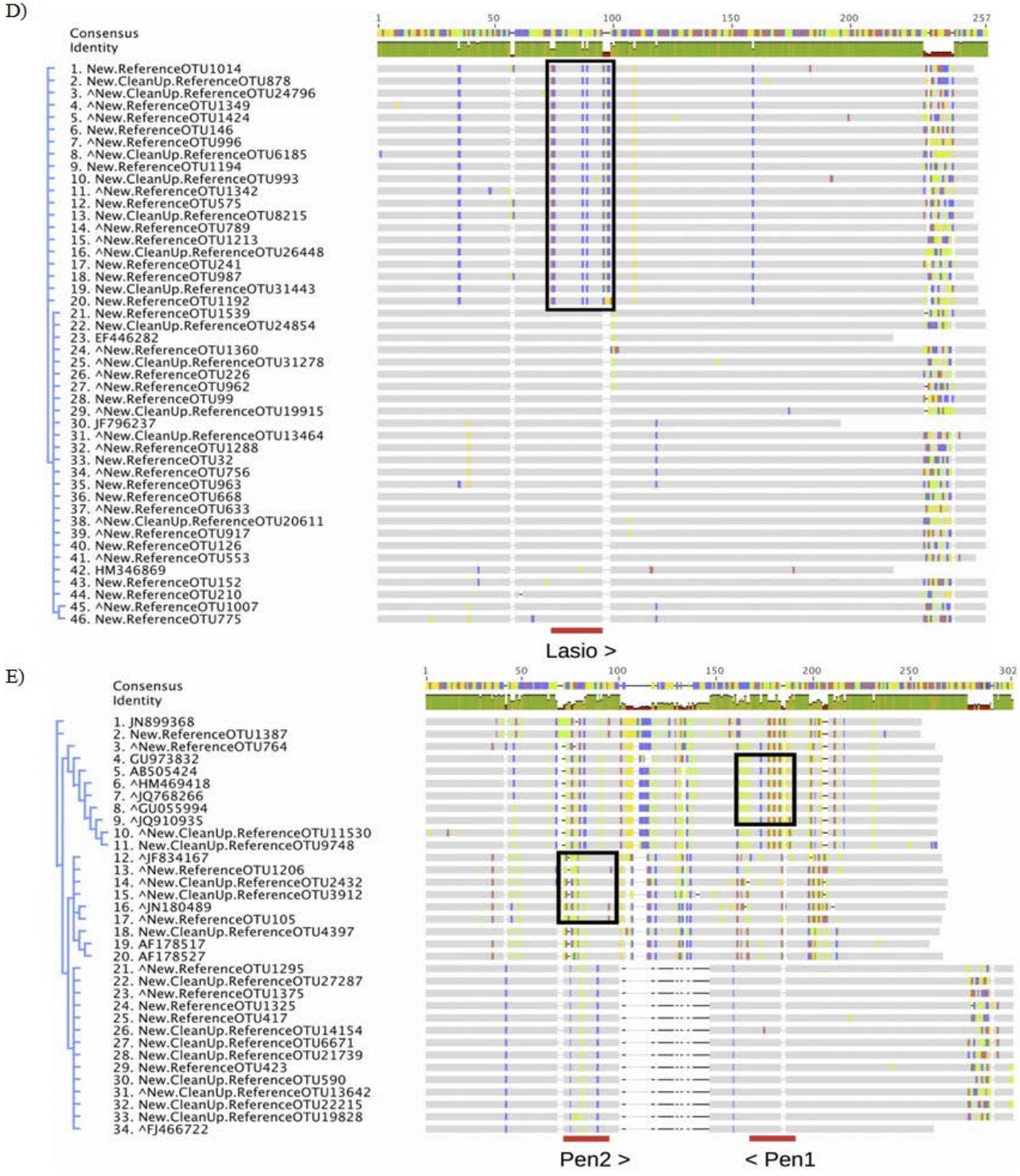


Figure 1. Alignment of OTU specific primers with target 16S and ITS sequences. Target regions are highlighted with black boxes. Primer sites and directionality are outlined with red bars in each figure. OTU names are the same as those in Debenport et al. (2015). Primers designed for clusters in A) *Chitinophaga*, B) *Aspergillus*, C) *Fusarium*, D) *Lasiodiplodia*, and E) *Penicillium* genera

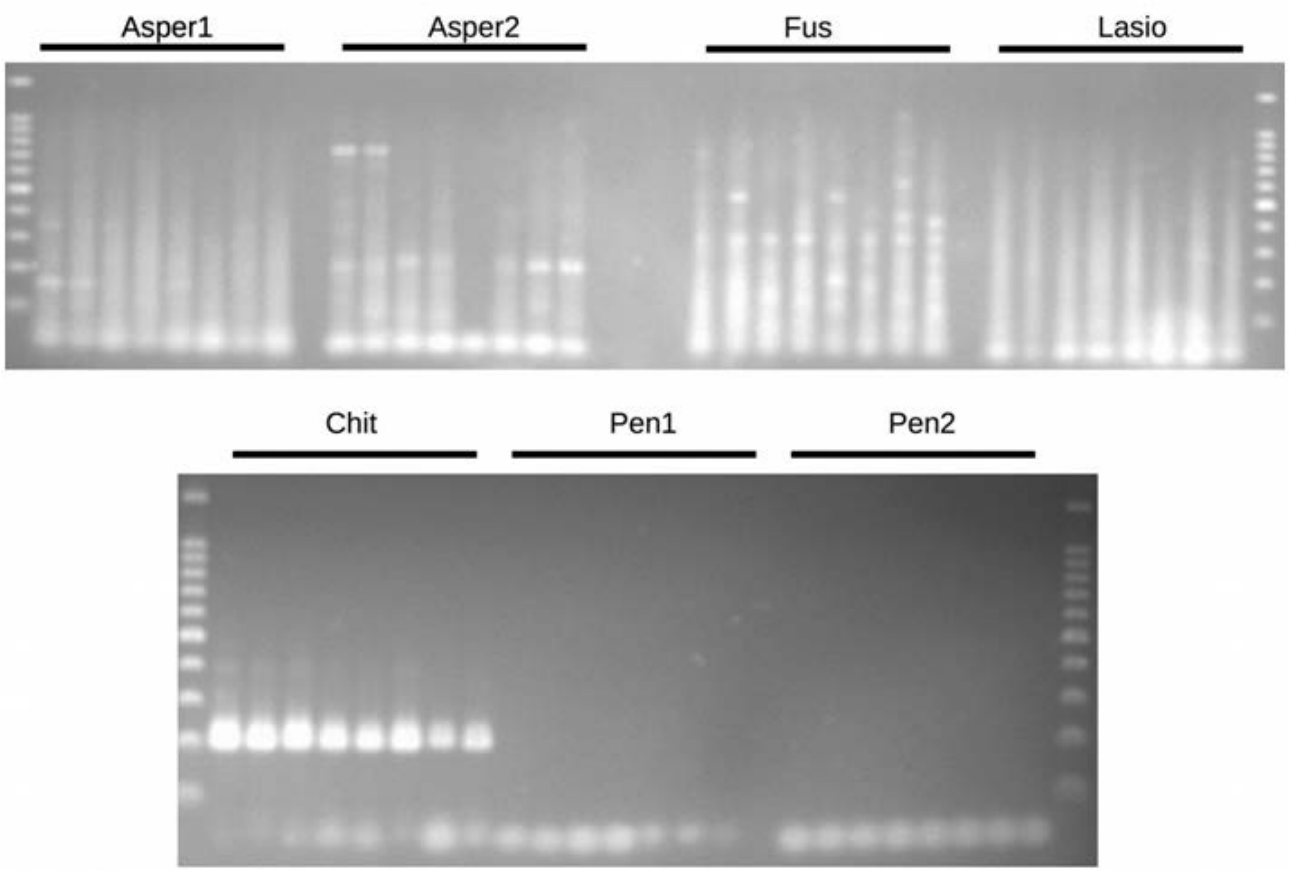


Figure 2. Specificity of OTU cluster-specific primer sets. Primer names are provided above each group of PCR products. Within each group of PCR products, the first four lanes use DNA template from intercropped millet samples and the last four lanes are from millet grown in bare soil. 100Bp ladder used

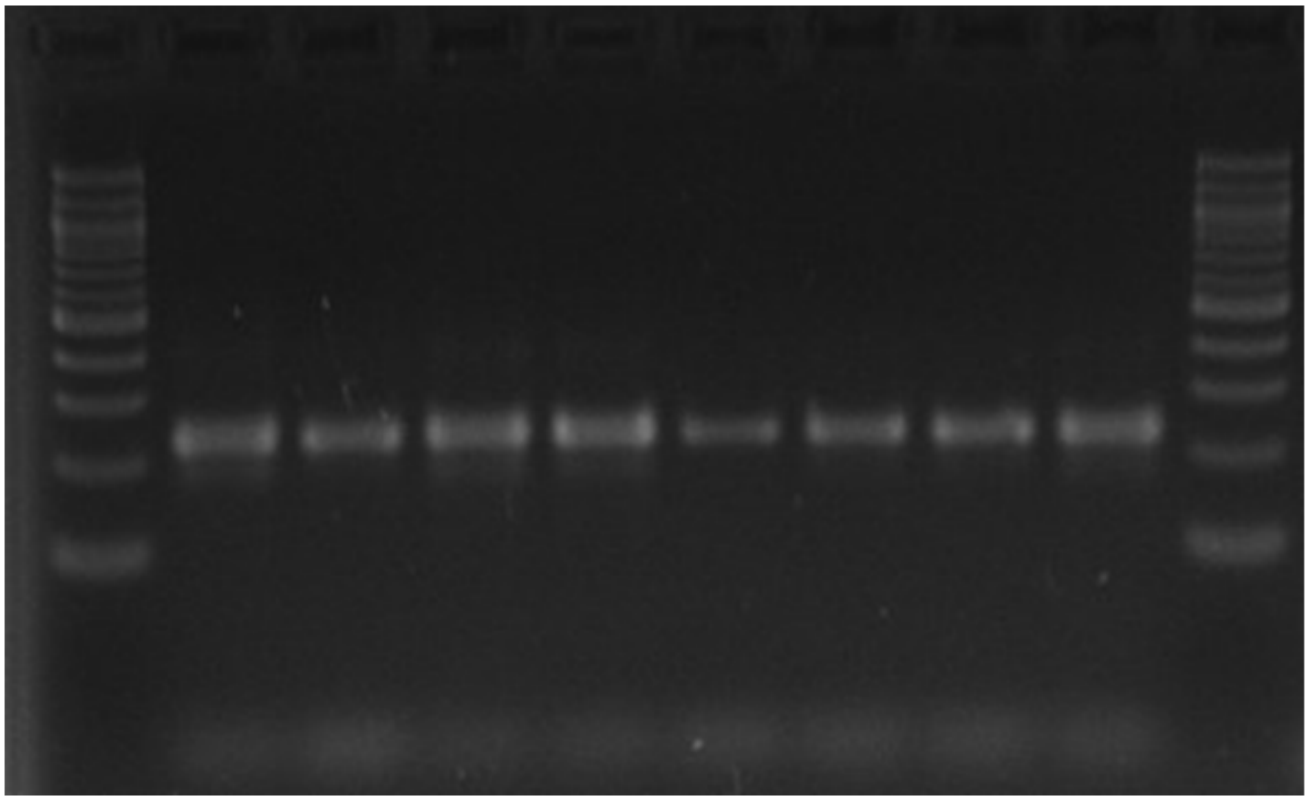


Figure 3. Specificity of *Chitinophaga* primer set. First four lanes use DNA template from intercropped millet samples. Last four lanes use DNA template from millet grown in bare soil. Bands from intercropped millet samples are 130% brighter on average. 100bp ladder used

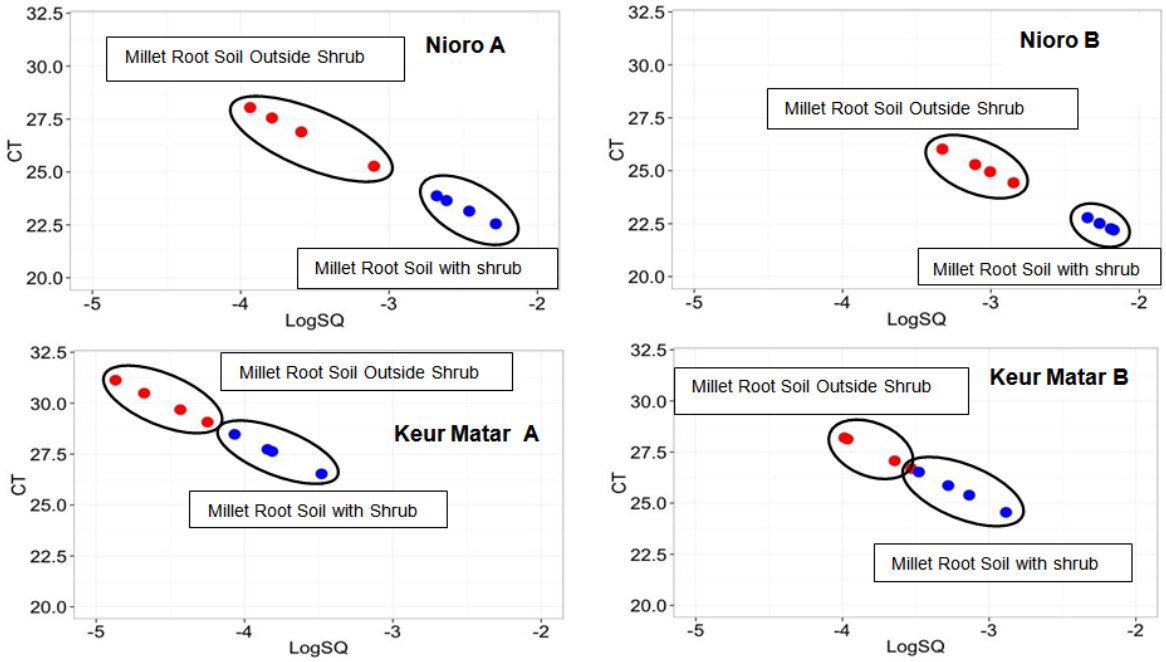


Figure 4. Quantification of *Chitinophaga* markers in millet root zone soils. Low CT (threshold cycle) and higher LogSQ (log starting DNA quantity) scores indicated higher abundance of molecular targets in sample template DNA. Intercropped millet samples are in blue, and millet grown in bare soil samples are in red

Table 2. Quantification and Fold Change Enrichment of *Chitinophaga* Targets. Fold Changes from Qpcr and Illumina Miseq Are Presented. Fold Changes Calculated as Enrichment in MR Samples Over MB Samples.

Expt. Type	Sample Location	Estimated Starting DNA Quantity (ng/µl)	Log2 Fold Change	MiSeq log2 Fold Change§
KM_A	MR	1.78×10^{-4} a†	2.47	6.55
	MB	3.19×10^{-5} b		
KM_B	MR	7.23×10^{-4} a	1.98	3.74
	MB	1.83×10^{-4} b		
N_A	MR	3.31×10^{-3} a	3.32	4.70
	MB	3.31×10^{-4} b		
N_B	MR	5.75×10^{-3} a	2.79	5.08
	MB	9.14×10^{-4} b		

† Values within an experiment followed by the same lower case letter are not significantly different at ($P < 0.05$).
§ Log 2 fold change of target *Chitinophaga* OTUs calculated from MiSeq analysis in Debenport et al. (2015).

4. Discussion

High-throughput sequencing of amplicon libraries allows for the presence and abundance of microbial groups to be compared between treatment or sample types at a fine level of resolution. While the cost of these techniques is steadily declining, rapid and inexpensive qPCR analysis is still a useful microbial research tool [21,22] (Kylmus et al., 2020; Larsen et al., 2020). Although all PCR-based techniques can be subject to amplification biases and sequencing data can result in artefacts [23] (Alteio et al., 2021), when used together, amplicon sequencing and qPCR can be used to overcome the limitations of each technique and strengthen the data in order to draw firmer conclusions. One such limitation to qPCR is that it is low-throughput [24] (Dreier et al., 2022). This can be ameliorated with the use of high-throughput amplicon sequencing to generate and test hypotheses which can then be further assessed via qPCR. Another advantage of using qPCR and amplicon sequencing in tandem is that the DNA inputs of the original amplicon sequencing step can be used for validation without the risk of introducing further biases [23] (Alteio et al., 2021).

A major drawback for qPCR is that the target must be a known sequence to allow for primer design and standard template generation. However, amplicon sequencing results can be used to generate specific primers for targeted qPCR amplification [9,25] (Saingam et al., 2018; Kuang et al., 2018). For example, alternate locations for cluster specific primers can be used, with a goal of including cluster specific bases at the 3' end of the primer sequence. An increase in the length of primer sequences can also capture more of the variable regions and reduce mismatches with non-target DNA. Primers can be designed with the opposite orientation that are specific to these clusters. While this will reduce the overall target size, amplicons will still be within the length requirements for qPCR studies. Creating primer pairs where both primers are specific to our clusters of interest will theoretically improve the specificity of our assays. In addition to the modifications to primer design, the PCR protocols for these primer sets can also be altered to improve reaction kinetics. With these modifications to the primer and PCR

design, it may be possible to improve the specificity of these primer sets and use them for additional qPCR testing. Of the strains highlighted in [11] (Debenport et al., (2015) as being differentially enriched, genus-specific primer generation was only successful for *Chitinophaga*. This highlights the difficulty in qPCR primer construction from HTS data but also provides a roadmap for future studies into genus-specific primer generation for qPCR.

The molecular markers identified in amplicon sequencing studies also offer a variety of opportunities for multiple new directions of research. A primer specific for a given OTU can be designed and enable rapid and inexpensive qPCR analysis for many samples for a variety of applications. For instance, molecular markers of plant growth promoting microorganism from field soil samples can be discovered using high throughput amplicon sequencing [11] (Debenport et al., 2015). This approach was used by [26] Pompanon et al., (2012) who used high throughput amplicon sequencing to identify prey organisms in predator guts, and the subsequent design of qPCR primers to track the consumption of prey at a large scale. Further, if an OTU is shown to be correlated to a particular phenomenon in a controlled study through an amplicon sequencing analysis, researchers may want to test that pattern across landscape levels, which is of major interest to our research group [11] (Debenport et al., 2015).

In the current study, use of genus-specific primers for qPCR validation corroborates amplicon sequencing results and corrects for primer biases and sequencing artefacts. Such biases may artificially inflate or deflate relative abundances of key taxa, making it difficult to make inferences about community composition downstream [23] [27] (Alteio et al., 2021; Westaway et al., 2018) as was observed when comparing the results of the current study to [11] Debenport et al (2015). For example, concerns were raised for conclusions from the amplicon sequencing results as there was a four-to-eight-fold increase in the abundance of *Chitinophaga* OTUs compared to this qPCR assay. The higher-magnitude differences found for *Chitinophaga* in the 2015 study could be explained by the lack of primer specificity to the target genus, a phenomenon that is nearly unavoidable in 16S rRNA amplicon sequencing [28] (Poretsky et al., 2014). Briefly, we suspect that the 806F - 515R primer set used in [11] Debenport et al (2015) allowed for the amplification of individuals with less than 97% sequence similarity with the *Chitinophaga* OTUs, which increased their magnitude of enrichment. This discrepancy highlights the need for independent validation of abundances with specifically designed qPCR primers.

Such independent validation of results is especially important for interpreting microbial ecological insights gained from soil environments, as the high complexity and spatial heterogeneity is common to soils [23] (Alteio et al., 2021). Furthermore, most soil microorganisms live in close association with microaggregates, and their connectivity and dispersal are highly limited by water availability creating greater community stochasticity than other 'well-mixed' environments [29] (Wilpiszeski et al., 2019). This high degree of heterogeneity, both biological and chemical, means that sampling "the same soil" is difficult to impossible, even in a small plot. Researchers in this field tend to take composite samples (i.e., taking multiple cores per experimental unit and homogenizing them, as in Debenport et al. (2015) [11], which may artificially alter relationships between the microbial community and the experimental treatment leading to misinterpretation or over-interpretation of results [23] (Alteio et al., 2021).

In the case of Debenport et al., (2015) [11], OTUs from the genus *Chitinophaga* (which contains known plant growth promoting rhizobacteria) were found to be four to eight-fold enriched +OSS plots. OSS has been linked to increased crop performance through several potential mechanisms, including the enrichment of a beneficial microbial community [11,30,31] (Bogie et al., 2018, Debenport et al., 2015, Mason et al., 2022). At first glance, it might be surmised that *Chitinophaga* was responsible for a significant proportion of crop improvement, given its PGPR status and its high enrichment in +OSS plots, and it has been proposed that this strain be used for further study into its use as a biocontrol agent [11] (Debenport et al 2015). Although members of *Chitinophaga* have been previously shown to have PGPR effects [32,33] (Rilling et al., 2018; Waim et al., 2020), the limitations of amplicon sequencing in soils make it so that further validation with genus-specific qPCR primers was necessary before advancing to inoculation studies. The pattern of *Chitinophaga* enrichment in +OSS plots was similar to findings in Debenport et al. (2015) [11]. The maintenance of the enrichment pattern with changes to the methods, demonstrated that qPCR can be used for independent validation of relative abundance patterns observed in amplicon sequencing experiments and may be used as evidence for further research into the use of *Chitinophaga* as a biocontrol agent.

Our study adds to a growing body of research [10] [6] (Yang et al., 2012; Kwak et al., 2018) seeking to compare the magnitude of differential abundance in OTUs between qPCR and meta-genomic studies to optimize both qPCR and PCR methodologies [5,34] (Andersen et al., 2017, Jian et al., 2020), combining HTS and qPCR techniques. qPCR primer design is limited to known sequences and those sequences that are generated during an amplicon sequencing experiment. Our study shows the potential to overcome this limitation by using NGS to design species-specific primers and ultimately be useful in improving microbial ecology research.

Acknowledgments

This work was funded through the National Science Foundation Partnerships for International Research and Education (PIRE) grant OISE-0968247. The authors thank Dr. Francesca Rotondo and Dr. Joe Vaughn for their assistance with qPCR techniques.

References

[1] Fierer N., Strickland M.S, Liptz, D., Bradford M.A., and Cleveland, C.C, "Global patterns in belowground communities," *Ecology Letters*, 12, 1238-1249. 2009.

[2] Berendsen R.L., Pieterse C.M.J., and Bakker P.A.H.M, "The rhizosphere microbiome and plant health," *Trends in Plant Science* 17, 478-486. 2012.

[3] Erlacher A., Cardinale M., Grosch R., Grube M., and Berg G,"The impact of the pathogen *Rhizoctonia solani* and its beneficial counterpart *Bacillus amyloliquefasciens* on the indigenous lettuce microbiome," *Frontiers in Microbiology*. 5, 1-8. 2014.

[4] Lee C.K, Herbold C.W., Polson, S,W., Wommack K.E., Williamson, S.J., McDonald, I.R., and Cary S.C, "Ground truthing next-gen sequencing for microbial ecology-biases and errors in community structure estimates from PCR amplicon pyrosequencing," *PLoS ONE* 7,e44224, 2012.

[5] Jian C., Luukkonen P., Yki-Järvinen H., Salonen A., and Korpela K, "Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling," PLoS ONE 15(1), e0227285, 2020.

[6] Kwak M., Kong H., Choi K., S-K., Ju Yeon Song, Lee J., Lee P.A., Choi S.Y., Seo M., Lee H.J., Jung E.J., Park H., Roy N., Kim H., Lee M.M., Kim J.f., Rubin E.M., and Lee S-W, "Rhizosphere microbiome structure alters to enable wilt resistance in tomato," Nature Biotechnology, 36, 1100-1109, 2018.

[7] Borneman J., Becker J.O., Bent E., Lanoil B., McSpadden Gardener B., Olatinwo R., Presley L., Scupham A.J., Valinsky L., and Yin B, "Identifying microorganisms involved in specific in situ functions: Experimental design considerations for rRNA gene-based population studies and sequence-selective PCR assays," in Hurst CJ, Crawford RL, Garland JL, Lipson D.A., Mills A.L., Stetzenbach L.D. (eds.), *Manual Environmental Microbiology (3rd Ed.)*. American Society for Microbiology, Ch 61, 748-757, 2007.

[8] Bustin S. A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., and Wittwer, C.T, "The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments," Clinical Chemistry. 55, (4) 611-622. 2011.

[9] Kuang J., Yan X., Genders A.J., Granata C., and Bishop D.J, "An overview of technical considerations when using quantitative real-time PCR analysis of gene expression in human exercise research," *PLoS ONE*, 13(5):e0196438. Published 2018 May 10. 2018.

[10] Yang, J.I., Reugger, P.M., McKenry, M.V., Becker, O., and Borneman J, "Correlations between root-associated microorganisms and peach replant disease symptoms in a California soil," *PLoS ONE*, 7, e46420, 2012.

[11] Debenport S.J., Assigbetse K., Bayala R., Chapuis-Lydie L., Dick, R.P., and McSpadden Gardener B.B, "Shifting populations in the root-zone microbiome of millet associated with enhanced crop productivity in the Sahel," *Applied and Environmental Microbiology* 8, 2841-2851, 2015.

[12] Bright, M., Diedhiou, I., Bayala, R., Assigbetse, K., Chapuis-Lardy, L., Ndour, Y., and Dick, R.P, "Long-term *Piliostigma reticulatum* intercropping in the Sahel: Crop productivity, carbon sequestration, nutrient cycling, and soil quality," *Agriculture, Ecosystems. and Environment*, 242, 9-22, 2017.

[13] Bright, M., Diedhiou, I., Bayala, R., Nathaniel Bogie, Chapuis-Lardy, L., Ghezzehei, T.A.,Chapuis-Lardy, L..A., Jourdan, C., Moucty Sambou, D., Badiane Ndour, Y., Cournac,L., and Dick, R.P, "An overlooked local resource: Shrub-intercropping for food production, drought resistance and ecosystem restoration in the Sahel," Agriculture, Ecosystems. and Environment. 319, 2021.

[14] 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. "QIIME allows analysis of high-throughput community sequencing data," *Nature Methods*, 7, 335-336, 2010.

[15] R Core Team. "R: a language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org, 2014.

[16] McMurdie P.J., Holmes S. "phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data," PLoS ONE. 8(4): e61217. 2013.

[17] Love M.I., Huber W., and Anders S, "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2," *Genome Biology*, 15, 550. 2014.

[18] Thompson J.D, Higgins D.G., and Gibson T.J, "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Research 22(22), 4673-4680, 1994.

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

[20] Gardes M., and Bruns T.D., "ITS primers with enhanced specificity for basidiomycetes – application to the identification of mycorrhizae and rusts," *Molecular Ecology* 2, 113-118, 1993.

[21] Klymus, K.E., Merkes, C.M., Allison, M.J., Goldberg, C., Helbin, C., Hunter, M., Jackson, C., Lance, R., Mangan, A., Monroe, E., Piagio, A., Stokdyk, J., Wilson, C., and Richter, C, "Reporting the limits of detection and quantification for environmental DNA assays, *Environmental DNA* 2, 271-282, 2020.

[22] Larsen, D.A., Wigginton, K.R, Tracking COVID-19 with wastewater, Biotechnology, 38, 1151-1153. 2020.

[23] Alteio, L.V., Séneca, J., Canarini, A., Angel,R., Jansa, J., Guseva, K., Kaiser, C., Richter, A. and Schmidt, H, "A critical perspective on interpreting amplicon sequencing data in soil ecological research," Soil Biology and Biochemistry, 160, 2021.

[24] Dreier M., Meola M., Berthoud H., Shani N., Wechsler D., and Junier P, "High-throughput qPCR and 16S rRNA gene amplicon sequencing as complementary methods for the investigation of the cheese microbiota,' *BMC Microbiology*. 7;22(1): 48. 2022.

[25] Saingam, P., Li, B., and Yan, T, "Use of amplicon sequencing to improve sensitivity in PCR-based detection of microbial pathogen in environmental samples," J. Microbiology Methods, 149: 73-79, 2018.

[26] Pompanon F., Deagle, B.E., Symondson, W.O.C., Brown, D.S., Jarman, S.N., and Taberlet, P," Who is eating what: Diet assessment using next generation sequencing," *Molecular Ecology* 21, 1931-1950, 2012.

[27] Westaway J.A.F., Huerlimann R., Miller C.M., Kandasamy Y., Norton R,, and Rudd D, "Methods for exploring the faecal microbiome of premature infants: a review," *Maternal Health, Neonatology and erinatology* i7(1), 11. 2018.

[28] Poretsky, R., Rodriguez-R. L.M., Luo C., Tsementzi D., Konstantinidis K.T, "Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics," *PLoS ONE*. 9(4): e93827.

[29] Wilpiszeski, R.L., Aufrecht, J.A., Retterer S.T., Sullivan, M.B., Graham, D.E., Pierce, E.M., Zablocki, O.D., Palumbo, A.V., and Elias, E.A, "Soil aggregate microbial communities: towards understanding microbiome interactions at biologically relevant scales," Applied and Environmental Microbiology, 85:e00324-19, 2019.

[30] Bogie N.A., Bayala, R., Diedhiou, I., Conklin M.H., Fogel, M.L., Dick R.P., and Ghezzehei, T.A, Hydraulic Redistribution by Native Sahelian Shrubs: Bioirrigation to Resist In-Season Drought. *Frontiers of Environmental Science*, 6. 2018.

[31] Mason, L.M., Delay, C.L., Debenport, S.J., Diedhiou, I., McSpadden Gardener, B., Assigbetse, K., Rich, V.I., and Dick, R.P, "Microbial community shifts in pearl millet root zone soils with Guiera senegalensis intercropping along a rainfall and soil type gradient in the Sahel" Soil Science Society of America Journal.

[32] Rilling J.I., Acuña J.J., Sadowsky M.J., and Jorquera M.A, "Putative Nitrogen-fixing bacteria associated with the rhizosphere and root endosphere of wheat plants grown in an Andisol from Southern Chile," Frontiers of Microbiology, Vol. 9. 2018.

[33] Wiam, A., Maged, S.M., and Heribert, H, "Desert Microbes for Boosting Sustainable Agriculture in Extreme Environments, Frontiers of Microbiology, 11, 2020.

[34] Andersen, S.C., Fachmann M.S.R., Kiil K., Møller Nielsen E., and Hoorfar J. "Gene-Based Pathogen Detection: Can We Use qPCR to Predict the Outcome of Diagnostic Metagenomics?" Genes 8(11): 332. 2017.