Updates to the 16S rRNA sequencing process at JGI

December 2015 *Adam R. Rivers*

Several improvements have been made to the amplicon sequencing process at JGI. This document describes changes to the 16S rRNA primer set used for the identification of bacteria and archaea in environmental samples. Since 2013, the V4 region of the 16S rRNA gene has been amplified using the modified 515f-806rB primer set (Caporaso *et al.*, 2012) and sequenced on Illumina MiSeq machines.

JGI is beginning to use to use a new primer set to amplify the V4-V5 region of the 16S rRNA gene. The new V4-V5 primer set (Parada *et al.*, 2015) was tested against the existing V4 primer set (Caporaso *et al.*, 2012) and a V3-V4 primer set (Takahashi *et al.*, 2014). For the comparison, 64 samples were used that spanned a range of environments including soil samples, plant endosphere and rhizosphere samples, marine samples from 0 to 200m depth and several mock communities including the extensively characterized 23 organism JGI mock community, and an insect gut mock metagenome.

Methods

All the primer sets contained Illumina linkers, a 12bp barcode index, a pad region, a 0,1,2 or 3 base pair long spacer and the sequence-specific primer. For the V4-V5 set and the V3-V4 set 8 forward and 8 reverse indices were used. For the V4 region 64 of JGI's standard reverse barcoded indexes and one forward index were used. For each primer set 64 samples were pooled and run on a 2 by 300bp Illimina MiSeq run. For each primer set, data were run through the Rolling Quality Control system at JGI and then annotated using the iTagger version 1.2 pipeline modified as appropriate with configurations and annotation databases for the new, longer read lengths.

New V4-V5 Primers

```
515F-Y 5′ GTGYCAGCMGCCGCGGTAA
926R   5′ CCGYCAATTYMTTTRAGTTT
```

(Parada *et al.*, 2015)

Questions

What is the benefit of the new primers?

More information. Sequencing improvements allow us to sequence two 300bp reads per amplicon. Previously we sequenced the ~292bp V4 region, but are now sequencing the longer ~412bp V4-V5 region of the 16S rRNA gene.

Less bias. All amplicon sequencing introduces bias, but the new 515F-Y-926R primer set reduces bias and can detect more environmentally important taxa including archaea.

Constancy. The Earth Microbiome Project has issued a protocol for V4-V5 sequencing and we think this primer set represents the emerging consensus sequence primer set. The new reads also entirely overlap with existing V4 reads.

Will V4-V5 work for my organism?

Probably, but we can't say for sure. Computational testing can be done using the Silva's Testprime tool using 1 or 2 mismatches but PCR is the best way of verifying compatibility. Please let us know if our primer set does not work for your organism.

Results

First, the proportion of reads making it through each phase of the iTagger annotation process was assessed. The V4 primer set produced the highest proportion of reads that that could be clustered into OTUs (94%) followed by the V4-V5 set (84%) and the V3-V4 set (66%) Fig. 1. Most loss occurred at the merging and extension phase. The length of the reads appears to be the cause of poorer merging performance. This is offset by the increased taxonomic resolution provided by the increased length, which can be seen in the percent of reads forming unique OTU clusters (23% for V4, 35% for V3-V4 and 38% for V4-V5). Fig. 1

Next we examined the taxonomic

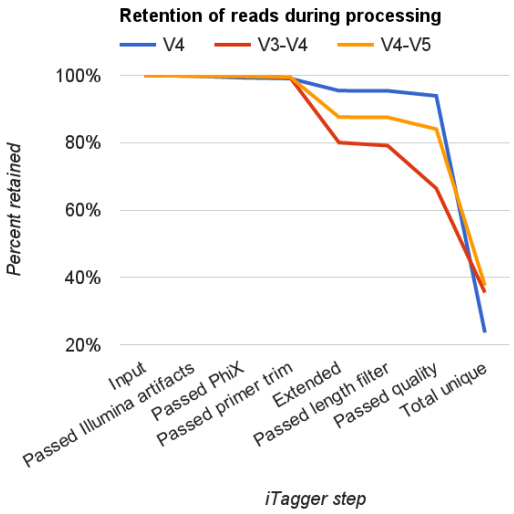

Figure 1

resolution of each OTU across the 64 samples. The longer regions resulted in more sequences being classified to lower taxonomic levels by the RDP classifier with a probability score of 0.8 or greater, Fig. 2. Gains were made at all taxonomic levels.

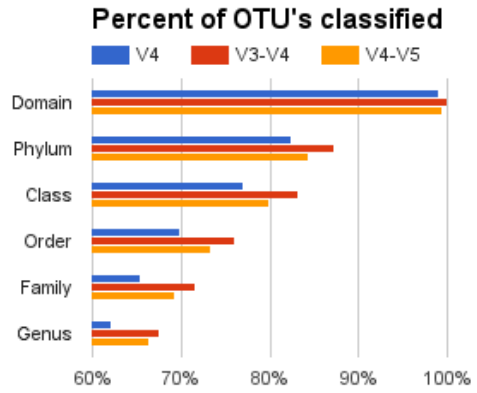

Figure 2

The accuracy of the three primer sets was assessed using the 23 organism JGI Mock Community. All three primer sets differed from expected 16S gene copy number estimated using the metagenome data. No one primer set was clearly the best at reproducing the community abundance of the mock metagenome. However the V4-V5 primer set captured every taxon in the mock metagenome while the V3-V4 primer set failed to amplify any of the archaea in the set (the halophilic Euryarchyota *Nantronococcus, Natronobacterium* and *Halovivax).* The V3-V4 and V4 primer sets did not detect the Actynomycete *Norardiopsis*. It is the relative abundance of OTUs across samples that is important for most applications, so we prioritized coverage over absolute quantification.

Selected Taxa	V3-V4	V4	V4-V5
Nitrosopumilus	No	No	Yes
Methanomicrobiales	No	Yes	Yes
Marine group II	No	No	Yes
Prochlorococcus	Yes	Yes	Yes
Pelagibacter	Yes	Yes	Yes
Pseudomonas	Yes	Yes	Yes
Rhodobacteraceae	Yes	Yes	Yes

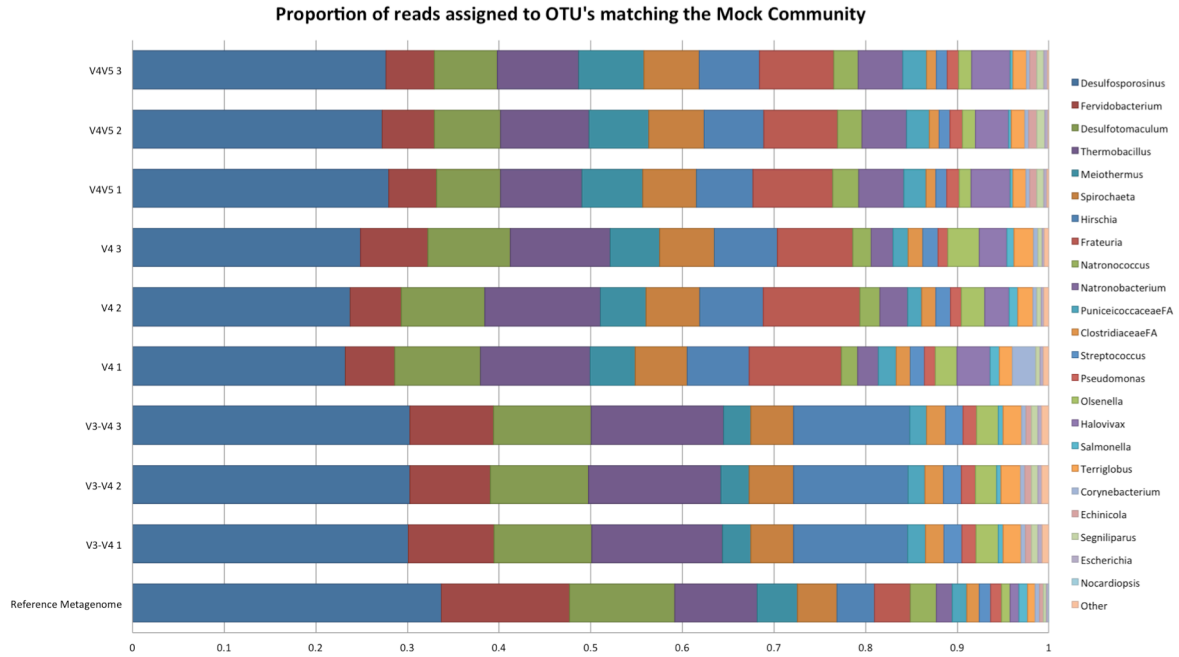

Figure 1

Another major goal in improving our primer sets was to improve coverage of key taxa that play important roles in ecosystems. The new V4-V5 primer set covers archaea better than the other two primer sets and has good coverage of other selected groups.

Conclusion

Ultimately, the decision to use the V4-V5 primer set came down to several factors. The increased length provides additional phylogenetic information without significantly compromising the number of reads that. The primer set does better with archaea than the two existing primer sets and it captures most taxonomic groups that users have expressed interest in. The selection of "universal" primers is always an exercise in compromise and there will likely be some taxa that are not represented as well. However, we feel that these V4-V5 primers will provide the best data to the most users.

References

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, *et al.* (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6**:1621–4. doi: 10.1038/ismej.2012.8

Parada A, Needham DM, Fuhrman JA. (2015). Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time-series and global field samples. *Environ Microbiol*. doi:10.1111/1462-2920.13023.

Takahashi S, Tomita J, Nishioka K, Hisada T, Nishijima M. (2014). Development of a prokaryotic universal primer for simultaneous analysis of Bacteria and Archaea using next-generation sequencing. *PLoS One* **9**:e105592. doi: 10.1371/journal.pone.0105592