A peer-reviewed version of this preprint was published in PeerJ on 8 August 2018.

View the peer-reviewed version (peerj.com/articles/5364), which is the preferred citable publication unless you specifically need to cite this preprint.

Nearing JT, Douglas GM, Comeau AM, Langille MGI. 2018. Denoising the Denoisers: an independent evaluation of microbiome sequence error-correction approaches. PeerJ 6:e5364 https://doi.org/10.7717/peerj.5364

PeerJ Preprints NOT PEER-REVIEWED

Denoising the Denoisers: An independent evaluation of microbiome sequence error-correction methods

Jacob T. Nearing[1], Gavin M. Douglas[1], André M. Comeau[2], Morgan G. I. Langille[1,2,3]

[1] Department of Microbiology & Immunology, Dalhousie University, Halifax, NS, Canada

[2] Integrated Microbiome Resource, Dalhousie University, Halifax, NS, Canada

[3] Department of Pharmacology, Dalhousie University, Halifax, NS, Canada

Corresponding Author:

Morgan Langille

Email address: morgan.langille@dal.ca

Abstract

High-depth sequencing of universal marker genes such as the 16S rRNA gene are a common strategy to profile microbial communities. Traditionally, sequence reads are clustered into operational taxonomic units (OTUs) at a defined identity threshold to avoid sequencing errors generating spurious taxonomic units. However, there have been numerous bioinformatic methods recently released that attempt to correct sequencing errors to determine real biological sequences at single nucleotide resolution by generating amplicon sequence variants (ASVs). As the microbiome field moves from OTUs to higher resolution ASVs, there is a need for an in-depth and unbiased comparison of these novel "denoising" methods. In this study, we conduct a thorough comparison of three of the most widely-used denoising methods on mock, soil, and host-associated communities. We tested three different methods - DADA2, UNOISE3, and Deblur - on four mock communities and found that, although they produced similar microbial compositions based on relative abundance, the methods identified vastly different numbers of ASVs. Our analysis of a soil dataset also showed that the three methods were consistent in their per-sample compositions, resulting in only minor differences based on weighted UniFrac distances. However, DADA2 tended to find more ASVs than the other two methods when analyzing both the real soil data and two other host-associated datasets, suggesting that it could be better at finding rare organisms. The three tested methods were significantly different in their run times, with UNOISE3 running greater than 1200 and 15 times faster than DADA2 and Deblur, respectively. Our results indicate that the choice of denoising method will depend on a researcher's individual importance for identifying rare ASVs, the availability of computational resources, and their willingness to support open-source or closed-source software.

46 **Introduction**

47

48 Microbiome studies often use an amplicon sequencing approach where a single genomic region,
49 such as part of the 16S rRNA gene (16S), is sequenced at a sufficient depth to provide relative
50 abundance profiles of the majority of the microbes in a sample. This sequencing approach is
51 often used to avoid the high cost of shotgun metagenomic sequencing or to avoid problems with
52 sequencing non-microbial DNA from host contamination. However, sequencing errors make it
53 difficult to distinguish biologically real nucleotide differences in 16S sequences from sequencing
54 artifacts. In the past, sequences were often clustered into operational taxonomic units (OTUs) at
55 a particular identity threshold (e.g. 97%) to avoid the problem of differentiating biological from
56 technical sequence variations. Recently, many new bioinformatic sequence "denoising" methods
57 have been developed to address this issue which provide improved species and strain resolution.
58 These methods differ in how they correct sequencing errors. DADA2 generates a parametric
59 error model that is trained on the entire sequencing run and then applies that model to correct and
60 collapse the sequence errors into what they call amplicon sequence variants (ASVs) (Callahan et
61 al., 2016). This method is advantageous as it builds unique error models for each sequencing run.
62 Deblur aligns sequences together into "sub-OTUs" and, based on the upper error rate bound
63 along with a constant probability of indels and the mean read error rate, removes predicted error-
64 derived reads from neighboring sequences (Amir et al., 2017). Deblur employs a sample-by-
65 sample method which reduces both memory requirements and computational demand. UNOISE3
66 uses a one-pass clustering strategy that does not depend on quality scores, but rather two
67 parameters with pre-set values that were curated by its author to generate "zero-radius OTUs"
68 (Edgar, 2016). The advantage of a one-pass clustering strategy is that it saves on the

computational time required to analyze the sequences in the provided study. Note that ASVs, sub-OTUs, and zero-radius OTUs are synonymous and the term ASV will be used henceforth. It is expected that denoising approaches provide improved resolution and they avoid having to make a choice between various OTU strategies which may result in differing results (Edgar, 2017). In addition, ASVs can be identified by their unique biological sequences instead of relying on per-study IDs, allowing for easier comparison across datasets (Callahan, McMurdie & Holmes, 2017).

Although there have been several bioinformatic comparisons of OTU-based approaches in the past (Allali et al., 2017; Plummer & Twin, 2015), a thorough third-party comparison of denoising methods has yet to be conducted. In this paper, we compare the strengths and weaknesses of DADA2, UNOISE3, and Deblur and assess their accuracy using several mock communities including both bacterial and fungal amplicons. In addition, we compare the results of the three methods on three previously-published real human, mouse, and soil datasets.

Material & Methods

Sequence Acquisition

The HMP mock community and the ZymoBIOMICS Microbial Community Standard (referred to as the Zymomock community) were sequenced by the Integrated Microbiome Resource at Dalhousie University using an Illumina MiSeq on separate sequencing runs, as previously described using the V4-V5 16S rRNA gene region (Comeau, Douglas, & Langille, 2017). Reads were then uploaded to the European Nucleotide Archive (ENA) under accession number PRJEB24409. The Extreme dataset (mock-12) originally presented in the DADA2 paper and the fungal ITS1 dataset (mock-9) were retrieved from the Mockrobiota project (Bokulich et al.,

2016). The Extreme dataset was sequenced using an Illumina MiSeq (Callahan et al., 2016) and the fungal mock community was sequenced using an Illumina HiSeq (Bokulich et al., 2016).

Filtering

All sample data were filtered using the Microbiome Helper filtering scripts (Comeau, Douglas, & Langille, 2017). In summary, primers were trimmed off all reads using Cutadapt (v 1.14) (Martin, 2011) and GNU Parallel (Tange, 2011). Primer-free sequences were then input into the dada2_filter.R script available in Microbiome Helper. This script takes in the maximum expected number of errors allowed as well as a truncation length. The HMP mock community and the Zymomock community were truncated to 270 and 210 base pairs for the forward and reverse read lengths. The single-end reads from the Extreme mock community and the fungal mock community were truncated to 80 base-pair lengths. The soil, mouse, and human-associated datasets were truncated to 270 and 210 base-pairs for the forward and reverse reads, respectively. The number of expected errors allowed were defined as three different filtering stringencies: 5 (low), 3 (medium), and 1 (high).

DADA2 Pipeline

The DADA2 method was run using scripts found in Microbiome Helper, which wraps the core algorithms of the DADA2 method (Callahan et al., 2016). Filtered reads were input into the wrapper script dada2_inference.R which runs the DADA2 inference algorithm. Once ASVs are determined, they are passed into DADA2's chimera-checking algorithm which was run using the wrapper script dada2_chimera_taxa.R to screen out chimeric sequences. The output objects

114 containing ASV sequences and abundances counts were then converted into BIOM table format
115 using convert_dada2_out.R. All DADA2 wrapper scripts were run with default settings.
116
117 **UNOISE3 Pipeline**
118 Filtered reads were input into USEARCH's (v 10) (Edgar, 2010) fastq_mergepairs command if
119 they were paired-end reads or concatenated together into one FASTQ if they were single-end
120 reads. Next, the single merged FASTQ was converted into a FASTA using the Microbiome
121 Helper script run_fastq_to_fasta.pl and then used as input for USEARCH's fastx_uniques
122 command which generated a FASTA containing all the unique sequences found in each sample.
123 Finally, the FASTA containing unique sequences was used as input into USEARCH's unoise3
124 (Edgar, 2016) command generating a BIOM table and representative ASVs that were used in
125 subsequent analyses. All USEARCH scripts were run with default settings.
126
127 **Deblur Pipeline**
128 Paired-end filtered reads were stitched together using the Microbiome Helper wrapper script
129 run_pear.pl which wraps the program PEAR (v 0.9.10) (Zhang et al., 2014). This step was
130 skipped for filtered single-end reads. Next, reads were renamed to match a format that was
131 compatible with QIIME2 (Caporaso et al., 2010b) and converted into a QIIME2 artifact. Samples
132 were then run through QIIME2's built-in deblur command using the 16S rRNA gene setting
133 which uses Greengenes 13_8 (DeSantis et al., 2006) for positive filtering. Fungal reads were run
134 using the "other" setting and the UNITE 10.10.2017 database (Kõljalg et al., 2013). Finally, the
135 representative ASV sequences and a BIOM table were exported from the QIIME2 artifact.
136

137 **Run Time and Memory Analysis**

138 Data from a blueberry field soil study (available under NCBI SRA PRJNA389786; Yurgel et al.,
139 2017) was filtered using the low-stringency filter and then individual samples were rarefied to
140 either 5000, 10000, 20000 or 30000 reads per sample. The different read-depth sets were then
141 run through the three denoising method pipelines and user time and maximum memory usage
142 was determined using the GNU time (v 1.7) command.

143

144 **ASV Analysis of Mock Communities**

145 ASVs were compared against the expected sequences provided with each of the mock
146 communities. This comparison was done using the command-line BLASTN (v 2.7.10) (Altschul
147 et al., 1990) tool and the number of full length 100% matches and 97% matches were
148 determined. All ASVs that did not match these criteria were then compared against the SILVA
149 16S rRNA gene database (v 128) (Pruesse et al., 2007) to find all 100% and 97% matches. Any
150 ASVs that did not match this database were then labeled as "Unmatched". To compare how
151 filtering of low abundance ASVs affected the type and amount of ASVs called by each method, a
152 0.1% minimum abundance filter was applied to each dataset and method.

153

154 **Abundance Data Analysis of Mock Communities**

155 For the HMP, zymomock and Extreme datasets all ASVs that matched at 97% identity or greater
156 with the provided expected sequences based on a BLASTN search were added to the abundance
157 of the corresponding matching taxa. Stacked bar charts of expected taxa relative abundances
158 were created using the ggplot2 (v 2.2.1) (Wickham, 2009) R (v 3.4.3) (R Development Core
159 Team, 2008) package and the cowplot (v 0.9.2) R package. The number of unique expected

160 sequences was determined by slicing out the amplified regions using a custom Python (v 3.6.1)
161 script (slice_amplified_region.py) from the expected sequences from each mock community.
162 Due to the incomplete nature of the expected sequences for the fungal mock community,
163 Unite database hits at 97% or greater to an expected sequence were considered as expected
164 ASVs. All other ASVs were classified as "Non-Reference" hits.
165
166 **Analysis of Real Datasets**
167 Data from the three real datasets: blueberry field soil (described above), stool from mice that
168 exercised plus controls (ENA accession PRJEB18615) (Lamoureux, Grandy, and Langille, 2017)
169 and the BISCUIT dataset of intestinal biopsies of pediatric Crohn's disease patients plus controls
170 (ENA accession PRJEB21933) (Douglas et al., 2018) were filtered using medium stringencies
171 for each denoising method and rarified to 5000, 3000 and 4259 reads, respectively. ASV
172 abundance tables outputted by all three methods were combined into a single table where each
173 biological sample was represented three times (once for each denoising method). ASVs not
174 called by a specific method were given an abundance of zero in their column (e.g. ASVs only
175 called by Deblur for sampleA were given zero abundances in the columns for DADA2's and
176 UNOISE3's outputs of sampleA). Representative sequences from each method were
177 concatenated into a single file and aligned using PyNast (Caporaso, Bittinger, et al., 2010a)
178 against the Greengenes alignment database (v13_8). A phylogenetic tree was then created using
179 the make_phylogeny.py script available in QIIME1 using the aligned sequences as input. A
180 weighted UniFrac distance matrix was generated using the beta_diversity.py command in
181 QIIME1. The distance matrix was then used to determine intra-sample distances between
182 methods, as well as to generate a principal coordinates analysis plot.

183 The Bray-Curtis distance matrix at the genus level was generated by assigning taxonomy
184 to the resulting ASV from each method using the RDP classifier (Cole et al., 2014) with the
185 assignTaxonomy function available in the DADA2 package and the rdp_train_set_16 database.
186 Distances were then generated using the summarize_taxa.py and beta_diversity.py commands in
187 QIIME1.
188
189 **Results**
190
191 **Total number of ASVs varies across methods**
192 We processed four different mock communities with the DADA2, UNOISE3, and Deblur
193 denoising pipelines to compare the resulting ASVs from each method. The number of called
194 ASVs varied between methods, but no method consistently called more ASVs. DADA2 called
195 the most ASVs in two communities (HMP: 42, Extreme: 74) and UNOISE3 called the most
196 ASVs in the other two communities (Zymomock: 43, Fungal: 37) under medium stringency
197 filtering (**Fig 1**). None of the methods output all expected sequences at 100% identity in any of
198 the mock communities that were processed and in all datasets at least one method output more
199 ASVs than expected sequences within the mock community. All three methods output at least
200 one ASV at 97% or greater identity from all organisms in the HMP mock community and the
201 Zymomock community (**Supp Table 1-4**). DADA2 output nine more ASVs with 97% or greater
202 identity matches to expected sequences in the Extreme dataset than the other two methods (**Supp**
203 **Table 2**). Five of the nine taxa that DADA2 called and the other methods did not call had
204 expected relative abundances of only 0.000427% (**Supp Table 2**). The other four taxa were also
205 in low expected abundances with one taxa being expected at 0.00427% (**Supp Table 2**). None of

206 the methods called any sequences that did not match either the expected sequences or the SILVA
207 database at 97% identity or greater for the Extreme dataset, which has previously been used for
208 validating both UNOISE2 and DADA2 (**Supp Table 2**).

209 Given that some of the above potential spurious ASVs would be removed by sequence
210 bleed-through (Illumina, 2017) or low abundance filters in typical workflows, we applied an
211 abundance cutoff filter of 0.1% abundance to the ASVs called by each method to see the effect
212 on the resulting abundances (**Supp Fig 1**). This resulted in all 10 unmatched ASVs (those that
213 did not match either the expected or SILVA by 97% or greater) called by DADA2 to be
214 discarded in the HMP community, but none of the four unmatched reads in UNOISE3 to be
215 discarded. A similar phenomenon was seen in the Zymomock community with all 12 of Deblur's
216 unmatched reads being discarded (along with one database hit) and UNOISE3 only discarding
217 one of 19 unmatched reads it called.

218 To determine how read quality filtering affects the number of ASVs called by each
219 pipeline, we ran all methods using two additional quality filtering stringencies, low and high (see
220 Methods). The different filter stringencies used made only small impacts on the numbers of
221 ASVs called by each method for the HMP, Extreme and fungal datasets. A difference of six
222 ASVs was the largest between the high and medium stringencies using the UNOISE3 method in
223 the HMP community (**Supp Table 5**). In the Zymomock community, the number of ASVs called
224 by DADA2 only varied by one for all three stringencies, but Deblur varied by as much as 12
225 ASVs and UNOISE3 varied by as much as 16 ASVs being outputted between the high and
226 medium filter stringencies (**Supp Table 5**).

227 We next wanted to see if these trends held in a real dataset, as the diversity of a mock
228 community is limited. In the soil dataset, DADA2 called 16609 ASVs, UNOISE3 called 11613

229 ASVs, and Deblur called 8270 ASVs after rarefaction (**Supp Fig 2a**). To determine how many of
230 these extra ASVs corresponded to new species, taxonomy was assigned used the RDP classifier
231 (see Methods). This showed that DADA2 called more classified taxa (413) than Deblur (315) or
232 UNOISE3 (360) (**Supp Fig 2b**). All of these extra taxa called by DADA2 were at abundances
233 less than 0.0006%. To confirm that DADA2 tended to call more ASVs in real datasets, the
234 denoising pipelines were also run on stool microbiome data from mice (exercise dataset) as well
235 as intestinal biopsy samples from pediatric patients (BISCUIT dataset). DADA2 called more
236 ASVs than the other two pipelines when run on each of these datasets, with DADA2 calling 727
237 more ASVs on average than Deblur and 532 more ASVs than UNOISE3 on average before
238 rarefaction (**Supp Table 6**).
239
240 **Methods are consistent in determining mock community composition**
241 Despite the different ASV counts between each method, the relative abundances of the expected
242 taxa are strikingly similar (**Fig. 2**). In both the HMP and zymomock datasets, only a small
243 portion of ASVs did not match the SILVA database by 97% identity or greater. In contrast,
244 UNOISE3 identified multiple (8 in HMP, 20 in Zymomock) sequences that summed together to
245 make up 2.5% and 4.6% of the relative abundance in the HMP and Zymomock communities,
246 respectively. None of the methods performed well at matching the expected abundance of the
247 Zymomock community or the fungal community. All three methods called over-abundances of
248 *Lactobacillus fermentum* in the Zymomock community. Similarly, all methods called non-
249 reference hits in greater than 10% abundance in the fungal community. Due to all three methods
250 producing similar results, this could suggest that either the mock compositions are not in the

251 expected proportions or that the three methods are similarly biased. Similar to above, the three
252 different filter stringencies resulted in similar relative abundance profiles (**Supp Fig 3**).
253

254 **Biological results from alternative methods are indistinguishable in real soil and host-**
255 **associated communities**

256 After comparing the relative abundances inferred by each method, we next investigated how
257 comparable the results between methods were for real 16S datasets. A soil dataset was chosen as
258 soil communities generally have high diversity (Fierer & Jackson, 2006) in direct contrast to the
259 limited diversity in mock communities. The intra-sample distances were compared between each
260 method using both weighted UniFrac and Bray-Curtis (based on genus-level taxonomy
261 assignment by the RDP classifier) metrics (**Fig 3**). All three methods had similarly small intra-
262 sample distances (~0.06) based on weighted UniFrac comparison (**Fig 3A**). Deblur-processed
263 samples had higher intra-sample Bray-Curtis distances (medians of 0.1707 vs. UNOISE3 and
264 0.1852 vs. DADA2) when compared with the other two methods (median 0.1193) (**Fig 3B**),
265 suggesting slightly higher agreement between DADA2 and UNOISE3 in comparison to Deblur.
266 This difference can be explained by a few outlying classifications because, in general, the
267 differences in relative abundances between the identified genera are close to 0 (**Supp Fig 6**).
268 DADA2 and UNOISE3 identified no genera in the soil dataset that differed by more than 1%
269 relative abundance, which contrasts with the comparisons of Deblur to DADA2 and UNOISE3 in
270 this dataset. Closer inspection of these outliers revealed that six of them were shared between the
271 comparisons of Deblur to DADA2 and UNOISE3. Two of the ASVs were assigned the same
272 class, Verrucomicrobia, but one of them was unclassified at the order level whereas the other
273 was placed in the Spartobacteria order (**Supp Fig 7B-C**). The abundances of these two classified

274 reads share an inverse relationship. The ASV classified at the order level was found in higher
275 abundance in DADA2 and NOISE3, but the unclassified ASV was found at higher abundance in
276 Deblur. Looking at another classification group, unclassified at the kingdom level, also shows
277 higher abundances found by Deblur than by DADA2 or UNOISE3 (**Supp Fig 7A**). Looking at
278 other taxa that have greater than 1% differences in abundance between Deblur and the other two
279 methods also revealed two similar classifications that differed at the order level and remained
280 unclassified at the family level. Deblur called more sequences for one of these classifications
281 whereas DADA2 and UNOISE3 called more for the other classification. These phylogenetically
282 close sequences could explain why the weighted UniFrac and Bray-Curtis distances show
283 different trends (**Supp Fig 7D-E**). Despite these differences, biological samples did indeed group
284 together regardless of the method used when visualized with either a principal coordinate
285 analysis (PCoA) plot (**Fig 3C**) or a non-metric multidimensional (NMDS) scaling plot (Fig 3D).
286 The same analysis was also done for the mouse exercise and BISCUIT datasets and we
287 found that in the mouse exercise dataset all three methods were equally similar for both weighted
288 UniFrac and Bray-Curtis metrics. On the other hand, in the BISCUIT dataset we found that again
289 Deblur was different in Bray-Curtis distances and also in weighted UniFrac distances. One large
290 driving force between these differences was the abundance of two different taxa, one in the
291 Lachnospiraceae family unclassified at the genus level and the other in the *Escherichia/Shigella*
292 genus. Deblur found higher abundances of the *Escherichia/Shigella* genus whereas DADA2 and
293 UNOISE3 found higher abundances of the Lachnospiraceae unclassified genus.
294
295 **Computational requirements are vastly different across methods**

296 Knowing that all three of these methods resulted in similar relative abundance profiles on mock
297 communities and small intra-sample distances on real 16S communities, we next investigated
298 how the run time and memory usage differed between the denoising methods. We found that
299 UNOISE3 (4.6 minutes) was 1272.52 times faster than DADA2 (5834.3 minutes) and 15.11
300 times faster than Deblur (69.3 minutes) at a total read count of 1,926,000 reads evenly distributed
301 across 103 samples (**Fig 4A**). Run times for all methods increased as the number of reads per
302 sample increased. Deblur used a static amount of memory (611 Mb) as reads per sample
303 increased, whereas in general the other two methods increased in memory usage as the number
304 of reads per sample increased with the exception of DADA2 run at 1,926,000 reads (**Fig 4B**).
305 Deblur used the smallest amount of memory at the maximum read count of 1,926,000 reads. We
306 found that DADA2 had the highest amount of memory usage (4071 Mb at 1,287,000 reads)
307 among the three methods. Interestingly, this usage was more than the amount used at the
308 maximum read count (3600 Mb). In addition, none of the runs exceeded the 4 Gb memory cap
309 on the 32-bit free academic version of USEARCH10.
310
311 **Discussion**
312
313 Besides specific differences in accuracy, there are other important aspects that need to be
314 considered when determining what method a researcher should use for their project. Both
315 DADA2 and UNOISE3 are suggested to be run in a pooled sample workflow, where all
316 sequences are pooled together during the denoising process (**Table 1**). Deblur, on the other hand,
317 runs its denoising process sample-by-sample. This approach helps lower Deblur's computational
318 requirements. Both DADA2 and Deblur are open source projects, whereas UNOISE3 is a closed-

source project which has a free 32-bit academic version with a 4 Gb memory cap and a full 64-bit version that costs between $885-1485 USD (**Table 1**). Another major difference is that the built-in Deblur function in QIIME2 has a positive filtering process. This default setting causes Deblur to discard reads that do not match with 88% identity to any sequences in the Greengenes database. Note the default database can be changed using the "other" version of the Deblur plugin in QIIME2, an important feature when working with fungal or eukaryotic data. It is also important to note that the stand-alone version of Deblur does not perform positive filtering by default, unlike the QIIME2 plugin which is the current version recommended by the authors. Currently, the functionality of both DADA2 and Deblur can be accessed through a graphical user interface as plugins in QIIME2, whereas UNOISE3 does not support a graphical user interface (**Table 1**).

During mock community data processing, no method consistently called more ASVs than another method. In addition, no method was able to call all expected sequences for each community at 100% identity. However, each method was able to detect every organism in the HMP community (note *S. aureus* and *S. epidermidis* are collapsed together as they have the same sequenced region) and the Zymomock community which in the end generated comparable relative abundance compositions to the expected amounts for the HMP community, but not the Zymomock community. In the Extreme dataset, all methods missed *P. buccalis*, *C. methylpentusum* and *P. sp._D13*. All three of these organisms had very low expected abundances (less than 0.00427%) which may explain why they were difficult to detect (**Supp Table 1**). Deblur and UNOISE3 both did not detect 9 of the 27 expected sequences in the Extreme dataset at 97% identity which were all detected by DADA2. Again, these nine organisms were at very low abundances (less than 0.05%). This difference in detection between DADA2 and the other

342 two methods suggests that it is better at detecting organisms that are very rare. Whether this
343 feature is truly advantageous is debatable, as many of these low-abundance organisms would be
344 removed by typical filtering cut-offs and/or contribute little to weighted beta-diversity metrics
345 such as the UniFrac measure.

346 To address the possibility of ASV abundance filtering, a minimum 0.1% abundance filter
347 was applied to the three different methods over all the datasets (**Supp Fig 1**). This filter cutoff
348 had a large effect on the number of unmatched ASVs called by DADA2 in the HMP mock
349 community (**Supp Fig 1a**) and the unmatched ASVs called by Deblur in the Zymomock
350 community (**Supp Fig 1c**), but had little effect on the number of ASVs called by UNOISE3 on
351 these communities. This cutoff had little to no effect on the fungal community (**Supp Fig 1b**).
352 One possibility for this occurrence is the difference in sequencing platforms as both the HMP
353 mock community and the Zymomock communities were sequenced on an Illumina MiSeq which
354 has an estimated sequence bleed-through rate of 0.1% (Illumina, 2017), whereas as the fungal
355 community was sequenced on an Illumina HiSeq. Overall, these results suggest that this filtering
356 practice may be useful when working on Illumina MiSeq data that has been processed using the
357 Deblur or DADA2 methods.

358 When testing the methods on a real soil dataset, DADA2 called significantly more ASVs
359 compared to the other two methods which is inconsistent with a previous report of UNOISE2
360 calling more ASVs than DADA2 in a soil sample (Edgar, 2016). This trend was confirmed on
361 the two other real 16S datasets used for validation. This discrepancy is most likely due to using a
362 different version of UNOISE, as UNOISE's chimera detection parameters were updated in its
363 latest iteration. Deblur, on the other hand, always called the least amount of ASVs among the
364 three methods on these three real datasets, although it was only slightly different than UNOISE3

365 on the human gut samples (difference of 44 ASVs before rarefaction). Deblur most likely called
366 the least amount of ASVs due to its positive filtering feature that discards reads not matching
367 with 88% identity to the Greengenes database. This feature is useful when dealing with well-
368 explored environments such as the gut, but could cause Deblur to miss many novel sequences in
369 less-characterized environments such as sediments (Karst et al., 2018). The rank-abundance
370 curve for the soil data also revealed that DADA2 called more rare taxa (**Supp Fig 2b**), similar to
371 what was seen in the Extreme mock community. This again indicates that DADA2 is better at
372 finding rare organisms within a sample.

373 Running the mock communities at different filtering stringencies had little effect on
374 microbial composition, which attests to the denoising capability of the three different methods
375 (**Supp Fig 3**). In general, allowing an increased number of expected errors resulted in more
376 sequences that did not match an expected sequence or a database, but did not have a large overall
377 effect. However, this finding was not true in all cases as DADA2 found more unmatched
378 sequences when the filter stringency was set to high in the HMP community (**Supp Table 5**).
379 However, this was not seen in the other three communities suggesting it is dataset-specific and
380 may not be a common occurrence.

381 The relative abundances determined for each study were similar to each other irrespective
382 of which method processed the data. This finding suggests that biological conclusions based on
383 microbial relative abundance data should be unaffected by the choice of denoising method. One
384 trend that was noticed in the relative abundance data was that UNOISE3 tended to call higher
385 abundances of non-reference ASVs. Interestingly, the lowest identity match for any of these
386 ASVs called in both the Zymomock and HMP mock communities by UNOISE3 was still found
387 at 90.4% identity to the SILVA 16S rRNA database and was classified as Gammaproteobacteria

388 by the RDP classifier using a 70% confidence threshold, suggesting it is a real biological
389 sequence that may have been introduced by contamination or sequencing bleed-through.
390 Importantly, these sequences were found at relatively low abundances and so had little impact on
391 the overall microbial compositions found in these mock communities (**Fig 2**).

392 The relative abundances determined within the Zymomock and fungal communities were
393 highly similar between methods, but markedly differed from the expected result. This finding
394 suggests that either the expected abundances of sequences from these communities may be
395 incorrect or all three methods are similarly biased. This non-agreement could also be due to steps
396 during the sequencing processes such as PCR amplification, which may be causing primer bias
397 (Aird et al., 2011) or the inclusion of contaminant organisms. In the case of the fungal
398 community, it is possible that none of these methods work well with ITS1 data which are more
399 variable than 16S data. Additional fungal mock communities should be analyzed in the future to
400 better explore this issue.

401 Benchmarking relative abundance profiles from different methods with mock
402 communities can be useful, however, they tend to lack the diversity that is found in many real
403 sample datasets. To address this issue, we compared resulting microbial compositions from each
404 method across three real datasets (mouse gut, human gut, and soil). Both weighted UniFrac and
405 Bray-Curtis distances between the same biological samples for each method were examined. In
406 both cases the weighted UniFrac and Bray-Curtis distances for all three datasets were small (less
407 than a median of 0.18) (**Fig 3a-b, Supp Figure 4a-b, 5a-b**). This complemented our previous
408 results, showing that each method had comparable microbial compositions for the mock
409 communities. Furthermore, plotting the samples on a PCoA or NMDS resulted in the same
410 biological samples from each pipeline grouping together (**Fig 3c-d**). This indicated that a similar

411 plot would be observed whether the researcher was using the Deblur, UNOISE3 or DADA2
412 method. Interestingly, Deblur did not agree with the DADA2 or UNOISE3 as much as they
413 agreed with each other on multiple occasions (**Fig 3b, Supp Fig 4a-b**). In the soil dataset,
414 differences in the Bray-Curtis distances, but not the weighted UniFrac distances, could be
415 explained by phylogenetically similar sequences being classified slightly differently, as well as
416 Deblur finding larger abundances of bacteria unclassified at the kingdom level (**Supp Fig 7**).
417 This result is interesting, as one of the main differences between Deblur and the other two
418 methods is its positive filtering feature, and so we expected this difference to drive Deblur into
419 finding less highly-unclassifiable, the opposite of what was seen in the soil dataset. Importantly,
420 this was not the case in the BISCUIT dataset where both the weighted UniFrac and Bray-Curtis
421 distances did not agree. In this dataset the differences between Deblur and the other two
422 denoising methods were driven by Deblur calling higher abundances of *Escherichia/Shigella* and
423 lower abundances of an unclassified genus in the Lachnospiraceae family. This indicated that
424 although in both cases Deblur did not line up with the other two methods it was for different
425 reasons. The mouse dataset did not show differences between Deblur and the other two methods.
426 A major difference between the three methods was their computational run time.
427 UNOISE3 was magnitudes faster than both DADA2 and Deblur. This is most likely due to both
428 the programming language that UNOISE3 is implemented in (C++), as well as its simple one-
429 pass denoising method. DADA2 was the slowest method and, although computation time could
430 be inconvenient for those with limited computational power, it did not reach times that were
431 impractical even when running almost 2 million total reads. Memory usage for each program
432 also did not reach impractical amounts when running close to 2 million reads, with DADA2
433 using a maximum amount of 1024 Mb of memory which is a reasonable amount for modern

434 computers. Memory usage by UNOISE3 did not come close to reaching the 4 Gb memory cap on
435 the 32-bit version, suggesting that this version can be used on most datasets.
436 In conclusion, all three methods are comparable when looking at their end results. The
437 main differences between the methods are the time taken to process data, as well as the number
438 of ASVs called. The number of ASVs called did not differ between methods in a consistent way
439 across mock communities, suggesting that determining species richness within low diverse
440 samples could be problematic. However, our analysis of real datasets showed that DADA2
441 consistently called more ASVs than the other two methods. More importantly, in the soil dataset
442 and in the Extreme dataset it was capable of finding more low-abundance organisms. In the end,
443 the choice of method did not play a large role in the microbial composition that was found for the
444 three mock communities. We believe this is a promising result, as it indicates that no matter the
445 choice of denoising method, the same biological signal will be observed. Our results also show
446 that the choice of denoising method will largely depend on the individual values of the
447 researcher that is using them, such as the importance of identifying rare organisms, the
448 availability of computational resources, and their willingness to support closed-source software.
449

450 **Acknowledgements**

451

452 We would like to thank members of the Langille lab for providing feedback and suggestions for
453 additional analysis during lab meetings. We would especially like to thank Karl Leuschen for
454 providing previous data on a comparison between Qiime OTU clustering and Dada2 denoising.
455
456

457 **References**

458

459 Aird, D., Ross, M. G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D. B.,
460 Nusbaum, C., & Gnirke, A. (2011). Analyzing and minimizing PCR amplification bias in
461 Illumina sequencing libraries. *Genome Biology*, *12*(2), R18. https://doi.org/10.1186/gb-
462 2011-12-2-r18

463 Allali, I., Arnold, J. W., Roach, J., Cadenas, M. B., Butz, N., Hassan, H. M., Koci, M., Ballou,
464 A., Mendoza, M., Ali, R., & Azcarate-Peril, M. A. (2017). A comparison of sequencing
465 platforms and bioinformatics pipelines for compositional analysis of the gut microbiome.
466 *BMC Microbiology*, *17*(1), 194. https://doi.org/10.1186/s12866-017-1101-8

467 Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local
468 alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.
469 https://doi.org/10.1016/S0022-2836(05)80360-2

470 Amir, A., McDonald, D., Navas-Molina, J. A., Kopylova, E., Morton, J. T., Zech Xu, Z.,
471 Kightley, E. P., Thompson, L. R., Hyde, E. R., Gonzalez, A., & Knight, R. (2017). Deblur
472 Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *mSystems*, *2*(2).
473 https://doi.org/10.1128/mSystems.00191-16

474 Bokulich, N. A., Rideout, J. R., Mercurio, W. G., Shiffer, A., Wolfe, B., Maurice, C. F., Dutton,
475 R. J., Turnbaugh, P. J., Knight, R., & Caporaso, J. G. (2016). mockrobiota: a Public
476 Resource for Microbiome Bioinformatics Benchmarking. *mSystems*, *1*(5).
477 https://doi.org/10.1128/mSystems.00062-16

478 Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace
479 operational taxonomic units in marker-gene data analysis. *The ISME Journal*, *11*(12), 2639–
480 2643. https://doi.org/10.1038/ismej.2017.119

481 Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P.
482 (2016). DADA2: High resolution sample inference from Illumina amplicon data. *Nature*
483 *Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

484 Caporaso, J. G., Bittinger, K., Bushman, F. D., DeSantis, T. Z., Andersen, G. L., & Knight, R.
485 (2010a). PyNAST: a flexible tool for aligning sequences to a template alignment.
486 *Bioinformatics*, *26*(2), 266–267. https://doi.org/10.1093/bioinformatics/btp636

487 Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K.,
488 Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights,
489 D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M.,
490 Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T.,
491 Zaneveld, J., & Knight, R. (2010b). QIIME allows analysis of high-throughput community
492 sequencing data. *Nature Methods*, *7*, 335. http://dx.doi.org/10.1038/nmeth.f.303

493 Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., Brown, C. T., Porras-
494 Alfaro, A., Kuske, C. R., & Tiedje, J. M. (2014). Ribosomal Database Project: data and
495 tools for high throughput rRNA analysis. *Nucleic Acids Research*, *42*(Database issue),
496 D633–D642. https://doi.org/10.1093/nar/gkt1244

497 Comeau, A. M., Douglas, G. M., & Langille, M. G. I. (2017). Microbiome Helper: a Custom and
498 Streamlined Workflow for Microbiome Research. *mSystems*, *2*(1).
499 https://doi.org/10.1128/mSystems.00127-16

500 DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T.,
501 Dalevi, D., Hu, P., & Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA
502 gene database and workbench compatible with ARB. *Applied and Environmental*
503 *Microbiology*, *72*(7), 5069–5072. https://doi.org/10.1128/AEM.03006-05

504 Douglas, G. M., Hansen, R., Jones, C. M. A., Dunn, K. A., Comeau, A. M., Bielawski, J. P.,
505 Tayler, R., El-Omar, E. M., Russell, R. K., Hold, G. L., Langille, M. G. I., & Van
506 Limbergen, J. (2018). Multi-omics differentially classify disease state and treatment
507 outcome in pediatric Crohn's disease. *Microbiome*, *6*(1), 13.
508 https://doi.org/10.1186/s40168-018-0398-3

509 Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST.
510 *Bioinformatics*, *26*(19), 2460–2461. http://dx.doi.org/10.1093/bioinformatics/btq461

511 Edgar, R. C. (2016). UNOISE2: improved error-correction for Illumina 16S and ITS amplicon
512 sequencing. *bioRxiv*. http://biorxiv.org/content/early/2016/10/15/081257.abstract

513 Edgar, R. C. (2017). Accuracy of microbial community diversity estimated by closed- and open-
514 reference OTUs. *PeerJ*, *5*, e3889. https://doi.org/10.7717/peerj.3889

515 Fierer, N., & Jackson, R. B. (2006). The diversity and biogeography of soil bacterial
516 communities. *Proceedings of the National Academy of Sciences of the United States of*
517 *America*, *103*(3), 626–631. https://doi.org/10.1073/pnas.0507535103

518 Illumina. (n.d.). Effects of Index Misassignment on Multiplexing and Downstream Analysis.
519 https://www.illumina.com/content/dam/illuminamarketing/documents/products/whitepapers
520 /index-hopping-white-paper-770-2017-004.pdf

521 Karst, S. M., Dueholm, M. S., McIlroy, S. J., Kirkegaard, R. H., Nielsen, P. H., & Albertsen, M.
522 (2018). Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene
523 sequences without primer bias. *Nature Biotechnology*. https://doi.org/10.1038/nbt.4045

524 Kõljalg, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F. S., Bahram, M., Bates, S.
525 T., Bruns, T. D., Bengtsson-Palme, J., Callaghan, T. M., Douglas, B., Drenkhan, T.,
526 Eberhardt, U., Dueñas, M., Grebenc, T., Griffith, G. W., Hartmann, M., Kirk, P. M.,
527 Kohout, P., Larsson, E., Lindahl, B. D., Lücking, R., Martín, M. P., Matheny, P. B.,
528 Nguyen, N. H., Niskanen, T., Oja, J., Peay, K. G., Peintner, U., Peterson, M., Põldmaa, K.,
529 Saag, L., Saar, I., Schüßler, A., Scott, J. A., Senés, C., Smith, M. E., Suija, A., Taylor, D.
530 L., Telleria, M. T., Weiss, M., & Larsson, K.-H. (2013). Towards a unified paradigm for
531 sequence-based identification of fungi. *Molecular Ecology*, *22*(21), 5271–5277.
532 https://doi.org/10.1111/mec.12481

533 Lamoureux, E. V, Grandy, S. A., & Langille, M. G. I. (2017). Moderate Exercise Has Limited
534 but Distinguishable Effects on the Mouse Microbiome. *mSystems*, *2*(4), e00006-17.
535 https://doi.org/10.1128/mSystems.00006-17

536 Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads.
537 *EMBnet.journal; Vol 17, No 1: Next Generation Sequencing Data Analysis*.
538 http://dx.doi.org/10.14806/ej.17.1.200

539 Plummer, E., & Twin, J. (2015). A Comparison of Three Bioinformatics Pipelines for the
540 Analysis of Preterm Gut Microbiota using 16S rRNA Gene Sequencing Data. *Journal of*
541 *Proteomics & Bioinformatics*, *8*(12). https://doi.org/doi: 10.4172/jpb.1000381

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glöckner, F. O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, *35*(21), 7188–7196. https://doi.org/10.1093/nar/gkm864

R Development Core Team. (2008). R: A Language and Environment for Statistical Computing. Vienna, Austria. http://www.r-project.org

Tange, O. (2011). GNU Parallel: the command-line power tool. *;login: The USENIX Magazine*, *36*(1), 42–47. https://doi.org/10.5281/zenodo.16303

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. http://ggplot2.org

Yurgel, S. N., Douglas, G. M., Comeau, E. M., Mammoliti, M., Dusault, A., Percival, D., & Langille, M. G. I. (2017). Variation in Bacterial and Eukaryotic Communities Associated with Natural and Managed Wild Blueberry Habitats. *Phytobiomes @BULLET*, *1*, 102–113. https://doi.org/10.1094/PBIOMES-03-17-0012-R

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, *30*(5), 614–620. https://doi.org/10.1093/bioinformatics/btt593

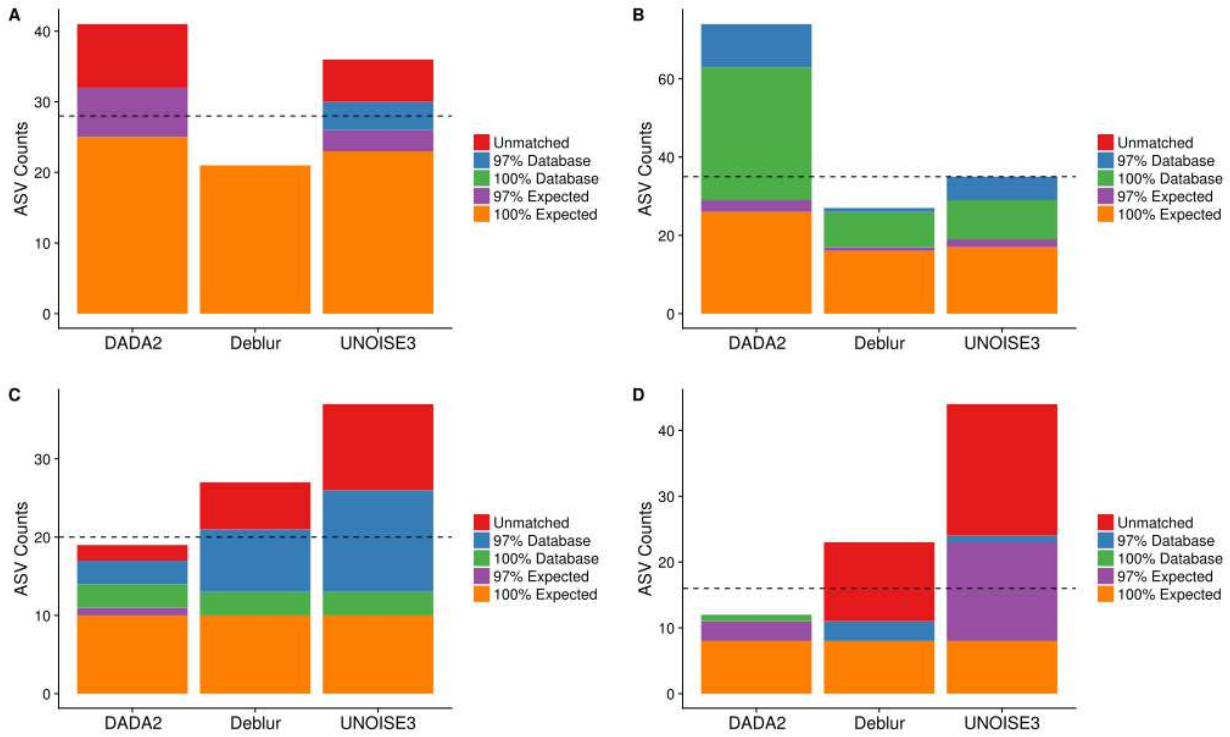


Figure 1: Total number of ASVs identified by each denoising method for four different mock communities. Amplicon sequence variants (ASVs) were compared to a database of full-length amplicon sequences for just the microbes supposedly in the community ("Expected") and against the full SILVA or ITS databases ("Database") using BLASTN at 97% and 100% identity cutoffs. "Unmatched" sequences did not match an expected sequence or the SILVA/ITS databases at 97% identity or greater. Dotted lines indicate the total number of ASVs expected, accounting for 16S copy variation within genomes. A) Human Microbiome Project mock community; B) Extreme dataset; C) Fungal ITS1 mock community; D) Zymomock community.

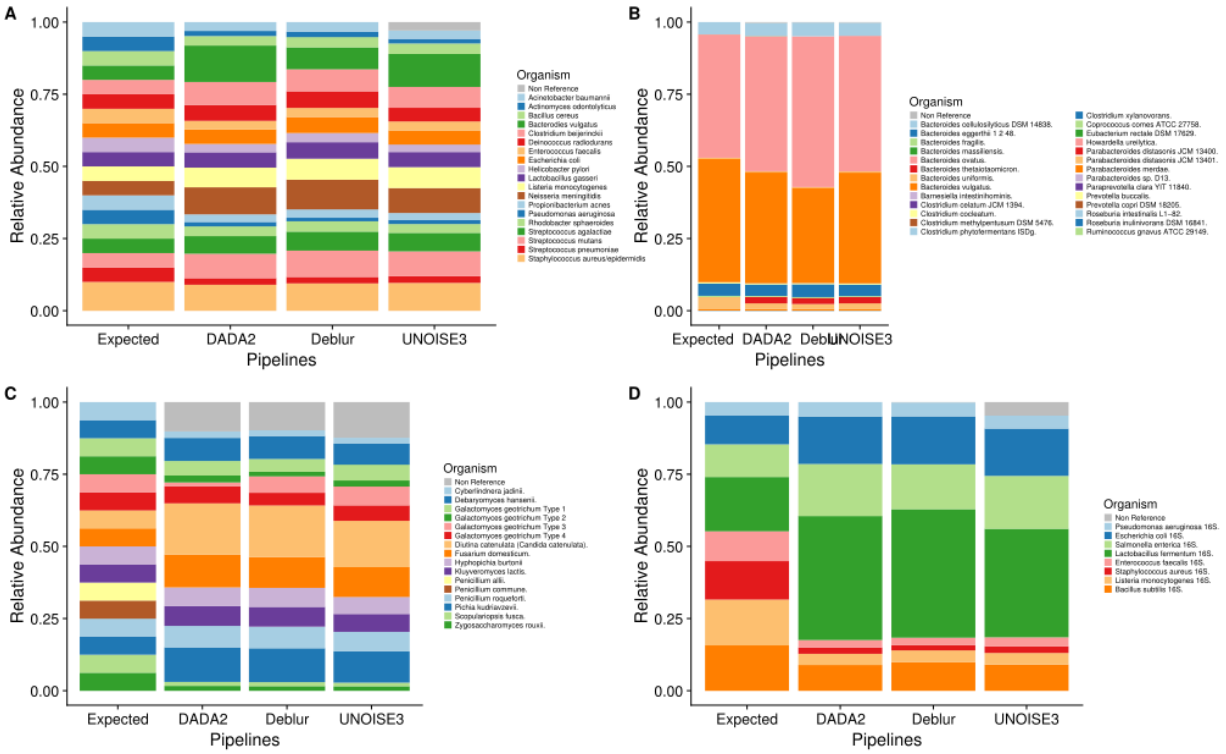


Figure 2: Relative abundances of taxa generated by each denoising method for four different mock communities. All ASVs that matched with expected sequences at 97% or greater identity were assigned taxonomy using a BLASTN search against the expected sequences provided for each the Extreme, Human Microbiome Project, and Zymomock mock communities. All ASVs that matched an expected species with 97% or greater identity to the UNITE database were classified as expected sequences for the fungal community. Non reference refers to the abundance of ASVs that did not match expected sequences with 97% or greater identity. A) Human Microbiome Project mock community; B) Extreme dataset - it is important to note that due to the low abundance of some organisms in the Extreme dataset they were not displayed in this figure; C) fungal ITS1 mock community; D) Zymomock community.

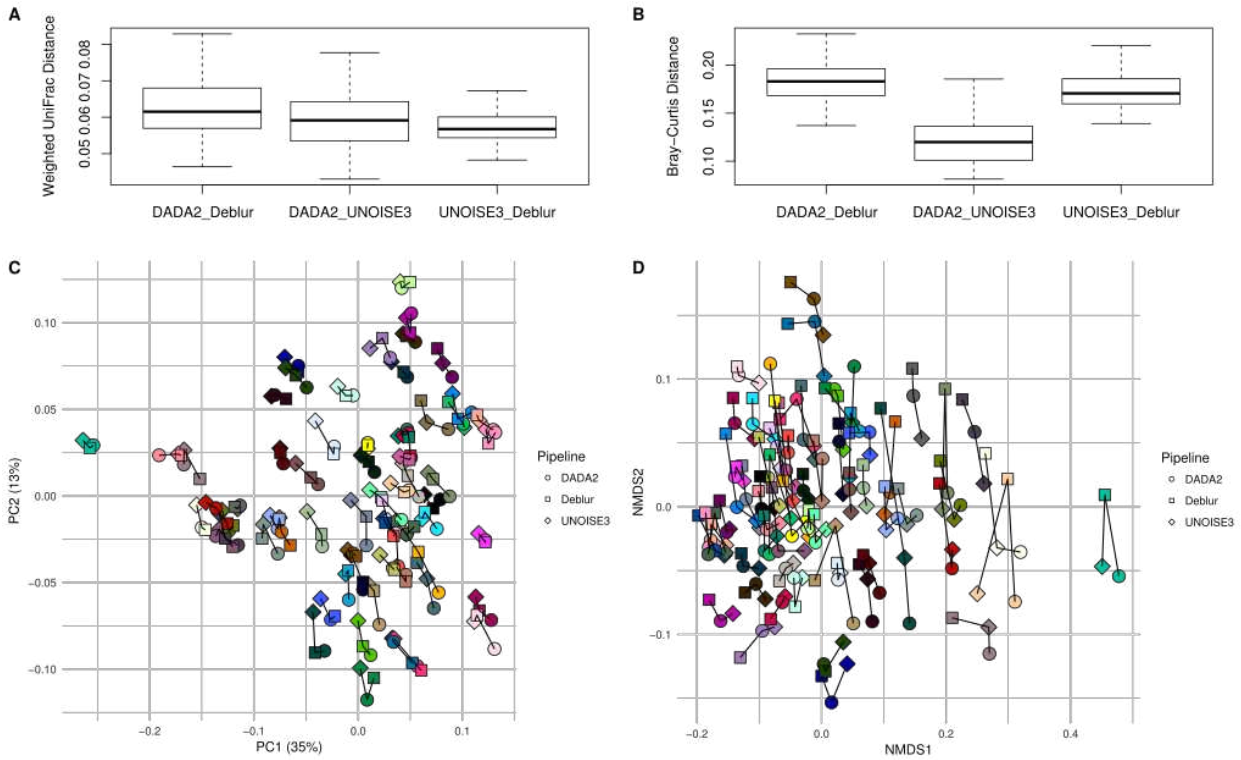


Figure 3: Intra-sample distances between denoising methods based on a real soil community. A) The weighted UniFrac distances between the same biological samples based on ASVs outputted by each of the different methods. B) The Bray-Curtis dissimilarity distances between the same biological samples based on genera outputted by the three methods after being classified with the RDP classifier. Deblur tends to be slightly more dissimilar when compared to the other two methods. C) Principal coordinates analysis of the weighted UniFrac distances of all the samples in the real soil dataset generated by each method. The three different profiles generated for each biological sample are colour-coded and are joined by an interconnecting line. D) Non-metric multidimensional scaling plot that displays the Bray-Curtis dissimilarity profiles of all the samples in the real soil dataset generated by each method. The three different profiles generated for each biological sample are colour-coded and are joined by an interconnecting line.

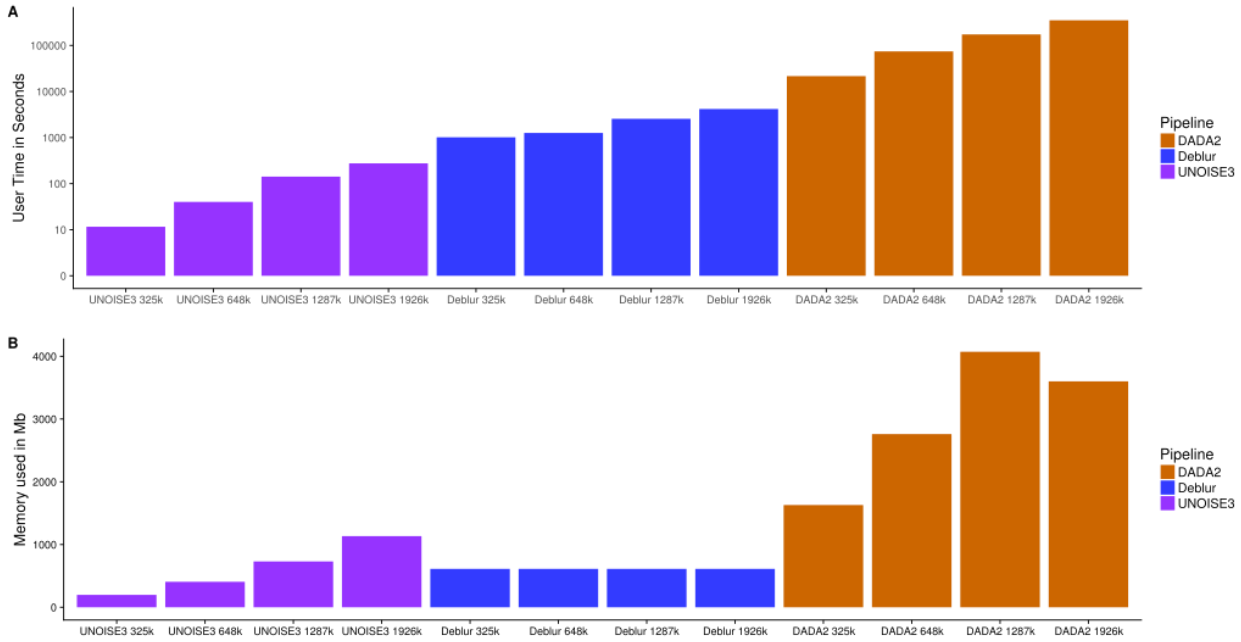


Figure 4: Run time and memory usage of each denoising method on a dataset of varying size. The time in seconds A) and memory in megabytes B) to run varying amounts of reads through the three different methods. Note time is on a $\log_{10}$ scale.

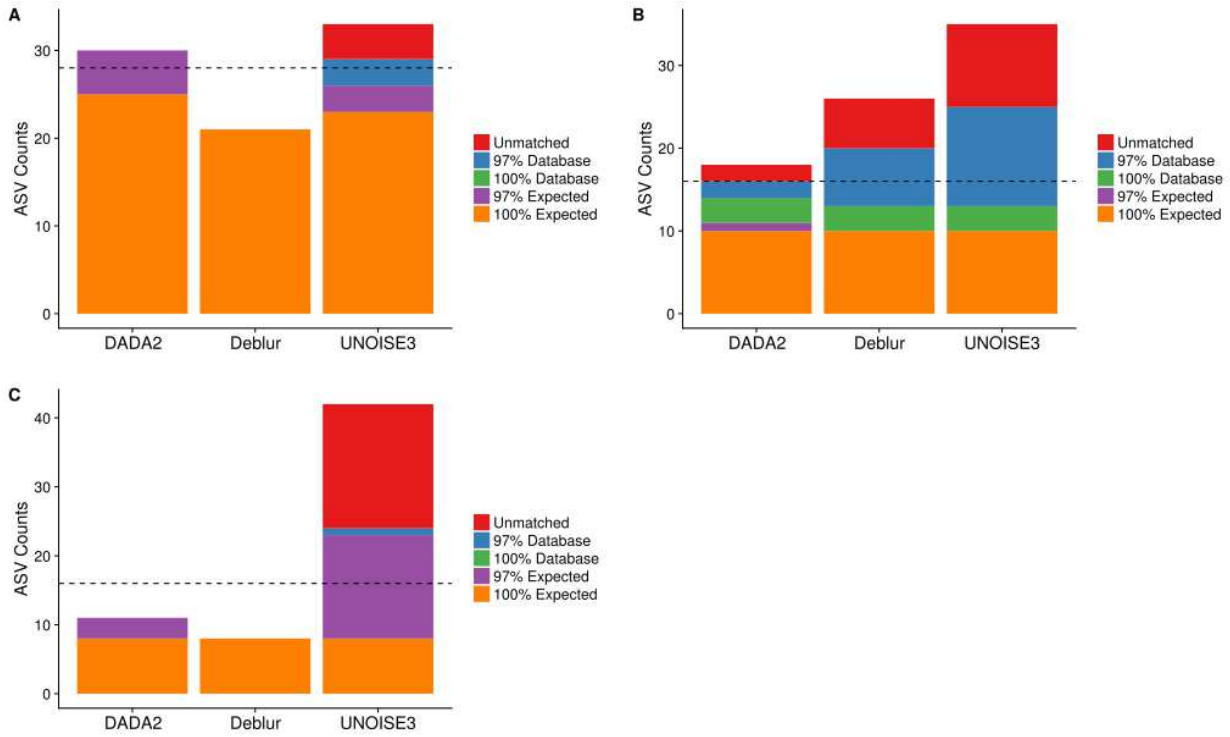
Table 1:

Qualitative Comparison of DADA2, Deblur, and UNOISE3

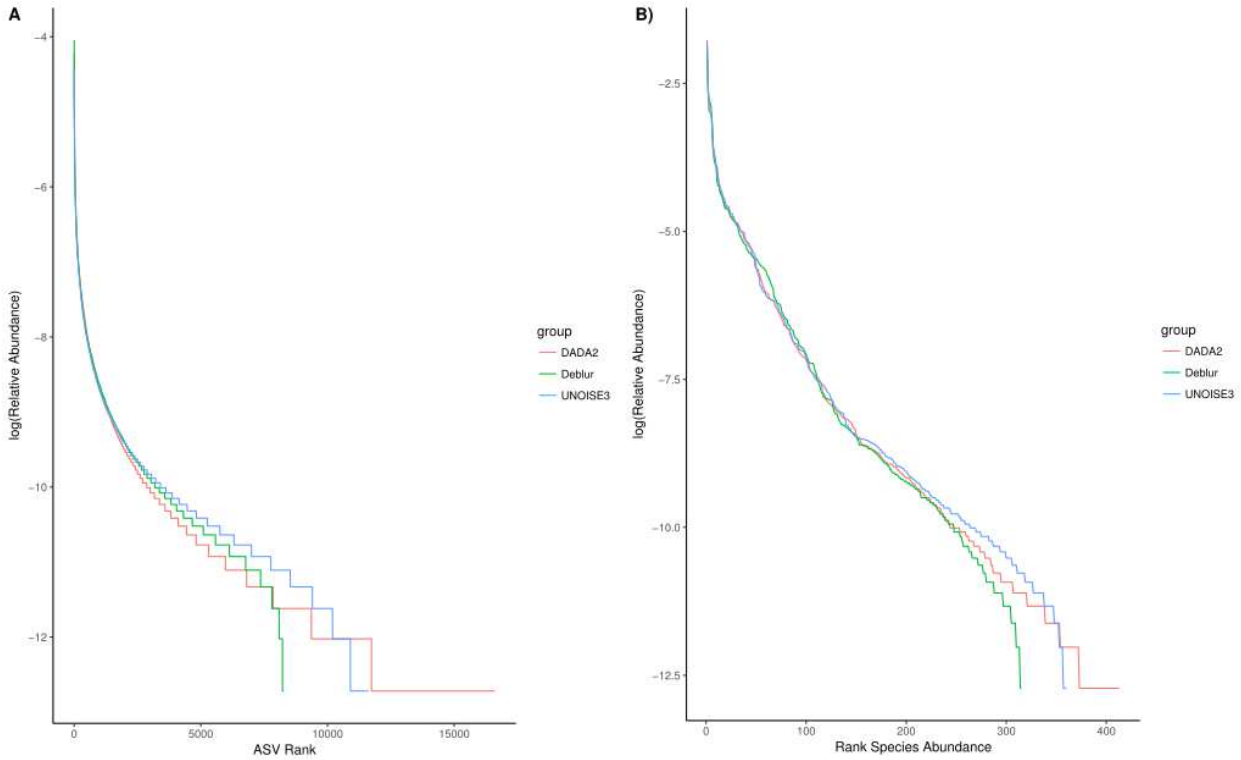
Pipeline	Implemented In	Open Source	*Pooled Sampling	**Positive Filtering	Version Tested	GUI via Qiime2	Publication Date
DADA2	R	Yes	Yes	No	1.6	Yes	April 13 2016
Deblur	Python	Yes	No	Yes	1.0.2	Yes	March 7, 2017
UNOISE3	C++	No	Yes	No	3	No	Oct 15, 2016

* When all sequences from all samples are denoised at the same time (in contrast to running each sample separately).

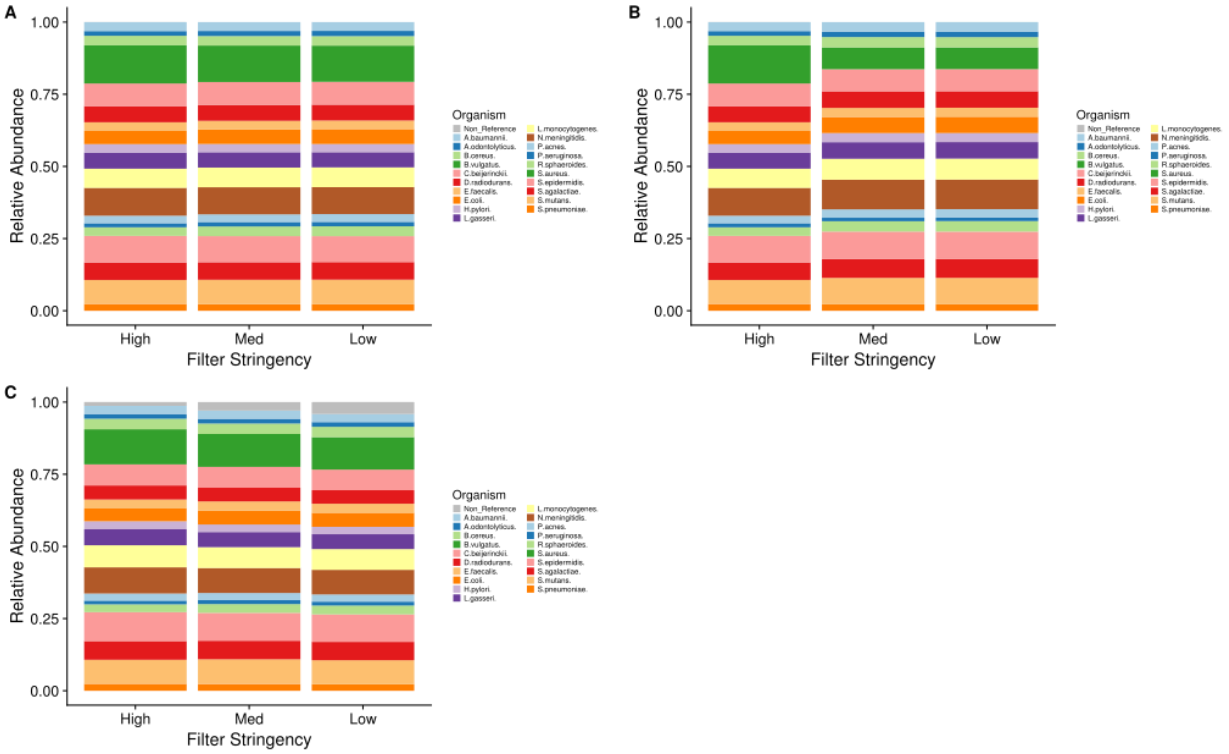
** Compares resulting ASVs to a database (Greengenes for Deblur) and discards reads if they do not match a certain identity threshold (88% for Deblur).



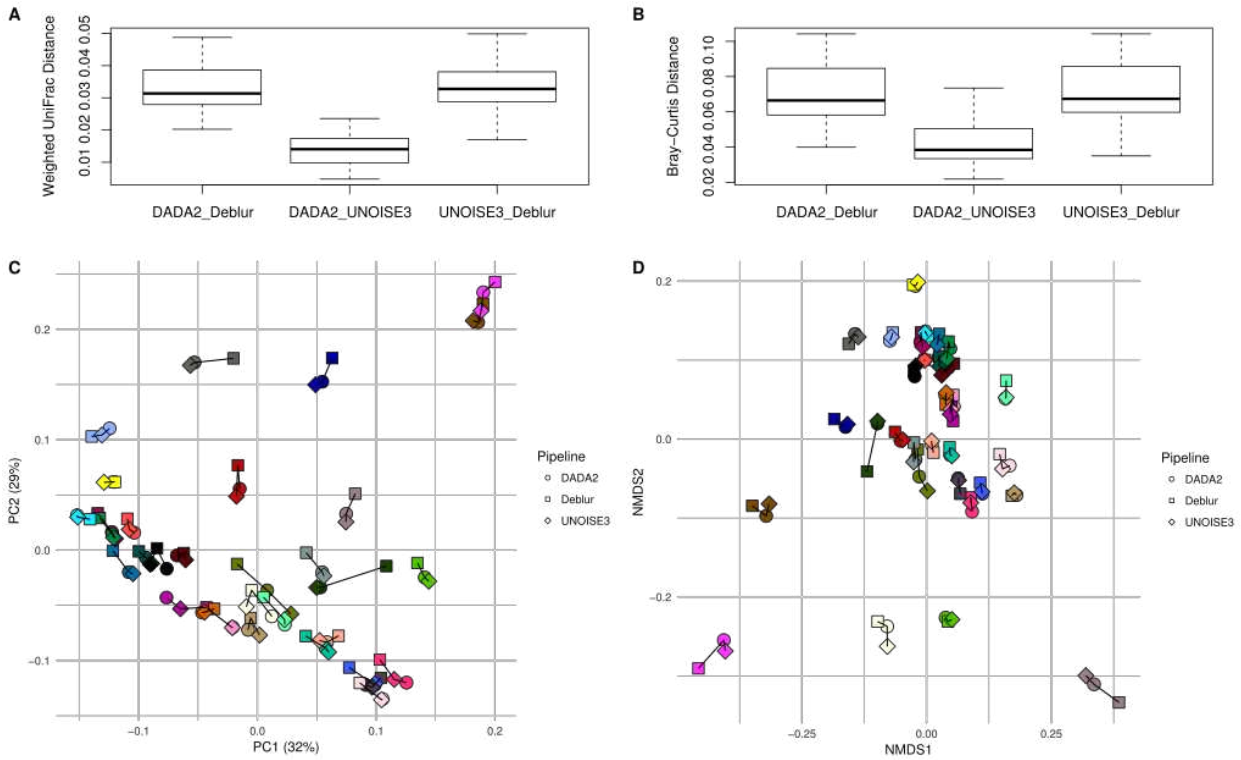
Supplemental Figure 1: Removal of low abundance ASVs removes many unmatched sequences from Deblur- and DADA2-generated ASVs. Amplicon sequence variants (ASVs) were run through an abundance filtering at 0.1% and then were compared to a database of full-length amplicon sequences for just the microbes supposedly in the community ("Expected") and against the full SILVA or ITS databases ("Database") using BLASTN at 97% and 100% identity cutoffs. "Unmatched" sequences did not match an expected sequence or the SILVA 16S rRNA gene database at 97% identity or greater. Dotted lines indicate the total number of ASVs expected, accounting for 16S gene-copy variations within genomes. A) Human Microbiome Project mock community; B) Extreme dataset; C) Fungal ITS1 mock community; D) Zymo mock community.



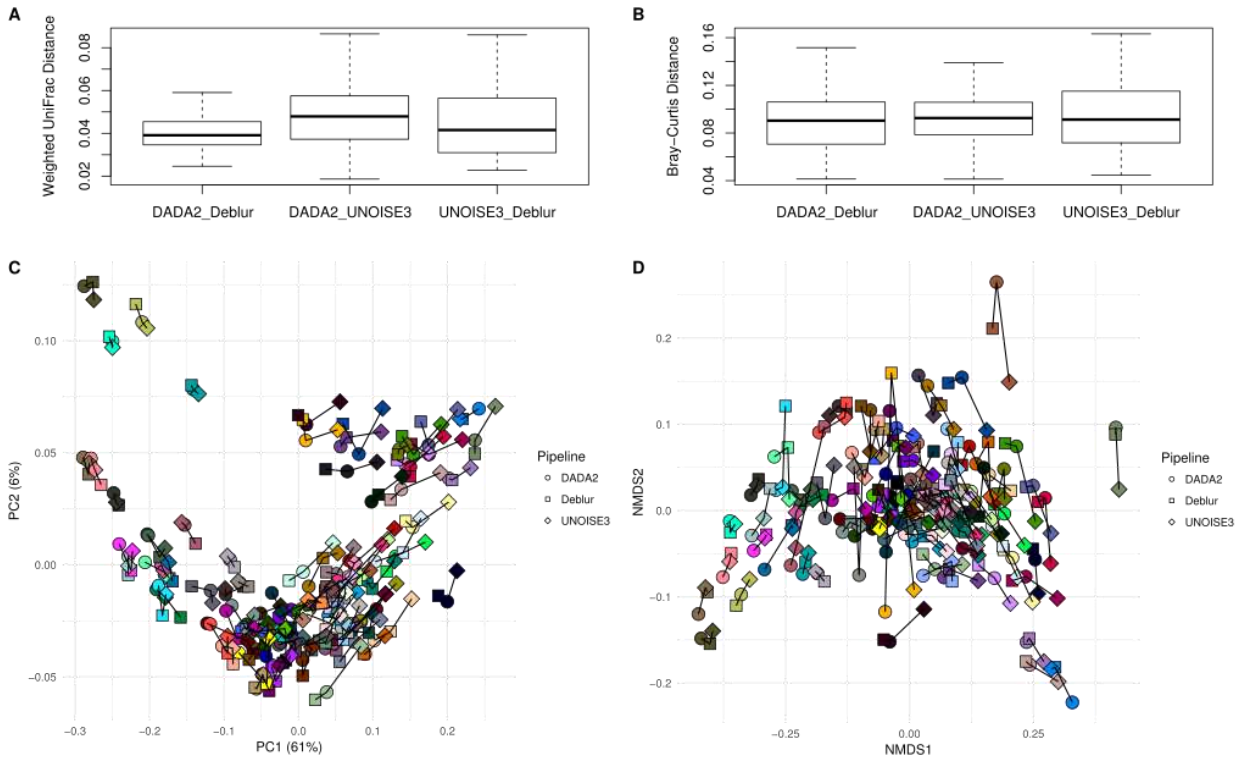
Supplemental Figure 2: DADA2 finds more rare organisms than Deblur or UNOISE3. Rank-abundance curves for ASVs (A) and classified species (B) generated from the soil dataset using the DADA2, Deblur and UNOISE3 methods. ASVs were classified using the RDP classifier against the Greengenes (13_8) database.



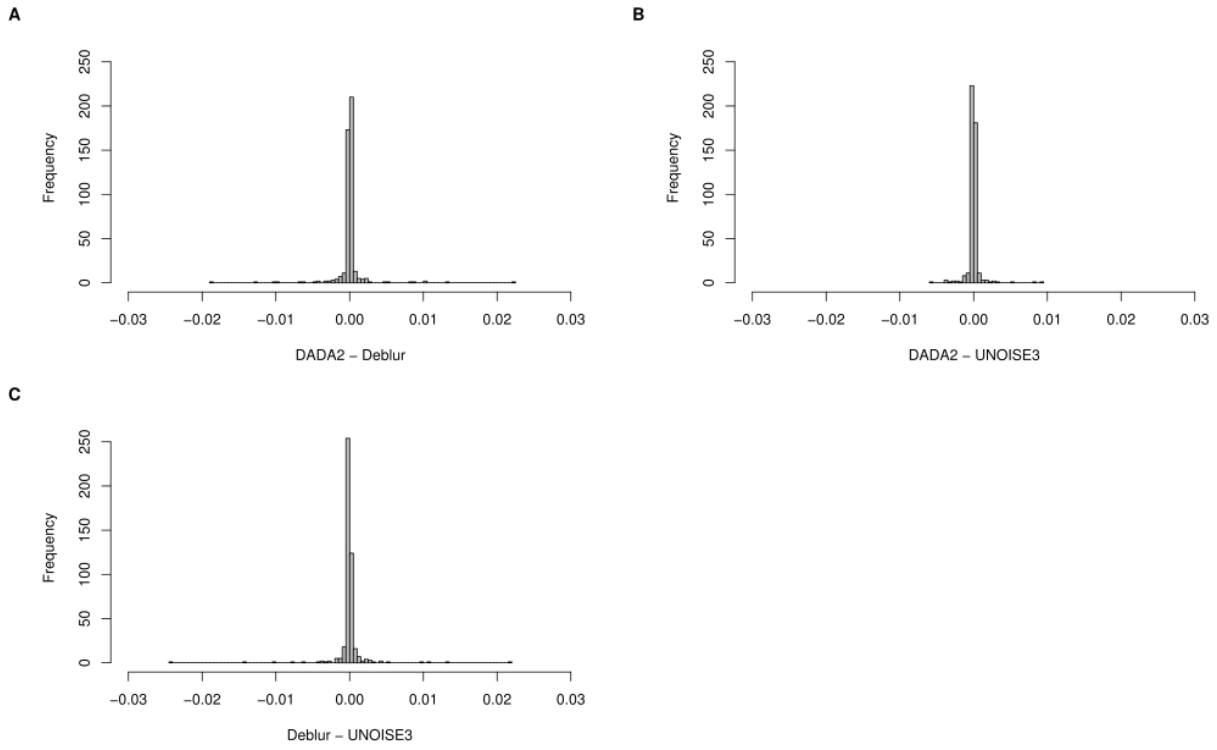
Supplemental Figure 3: Filter stringency does not affect relative abundance data drastically. The Human Microbiome Project mock community was run using DADA2, UNOISE3, and Deblur at varying stringency filters (low, medium and high). Resulting relative abundance profiles are shown for A) DADA2, B) Deblur and C) UNOISE3.



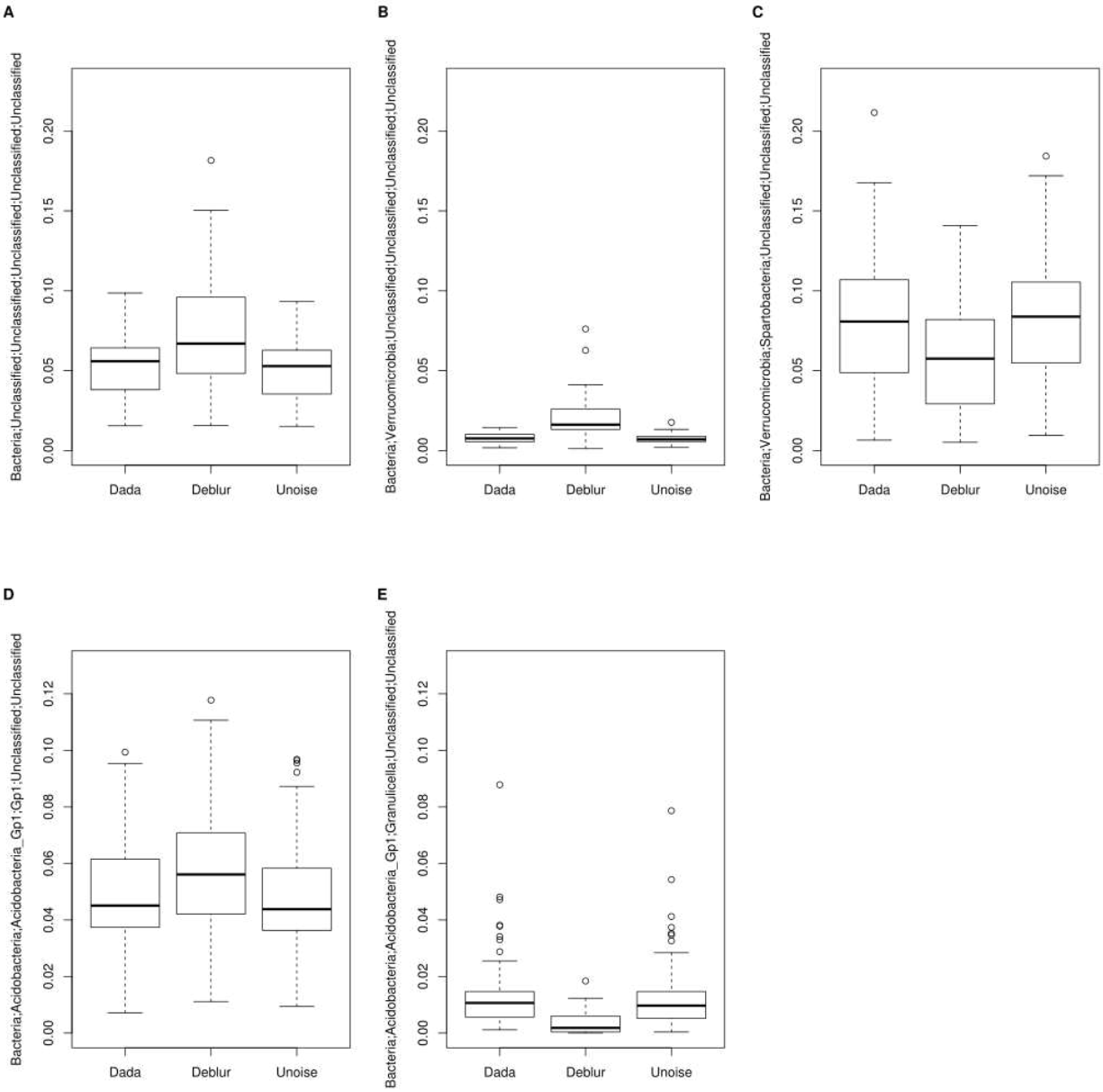
Supplemental Figure 4: Intra-sample distances between methods based on intestinal biopsy samples from pediatric Crohn's disease patients and controls. A) The weighted UniFrac distances between the same biological samples based on ASVs outputted by each of the different methods. B) The Bray-Curtis dissimilarity distance between the same biological samples based on genera outputted by the three methods after being classified with the RDP classifier. C) Principal coordinates analysis of the weighted UniFrac distances of all the samples in the real soil dataset generated by each method. The three different profiles generated for each biological sample are colour-coded and are joined by an interconnecting line. D) Non-metric multidimensional scaling plot that displays the Bray-Curtis dissimilarity profiles of all the samples in the real soil dataset generated by each method. The three different profiles generated for each biological sample are colour-coded and are joined by an interconnecting line.



Supplemental Figure 5: Intra-sample distances between methods based on mouse exercise associated fecal samples. A) The weighted UniFrac distances between the same biological sample based on ASVs outputted by each of the different methods. B) The Bray-Curtis dissimilarity distance between the same biological samples based on genera outputted by the three methods after being classified with the RDP classifier. C) Principal coordinates analysis of the weighted UniFrac distances of all the samples in the real soil dataset generated by each method. The three different profiles generated for each biological sample are colour-coded and are joined by an interconnecting line. D) Non-metric multidimensional scaling plot that displays the Bray-Curtis dissimilarity profiles of all the samples in the real soil dataset generated by each method. The three different profiles generated for each biological sample are colour-coded and are joined by an interconnecting line.



Supplemental Figure 6: There are outlier genera that drastically differ in relative abundance between Deblur and the other denoising methods. ASVs were classified using the RDP classifier against the Greengenes (13_8) database. Relative abundances of each genus were than compared between methods and differences were plotted in a histogram. A) Relative abundance differences by genus between DADA2 and Deblur. B) Relative abundance differences by genus between DADA2 and UNOISE3. C) Relative abundance differences by genus between Deblur and UNOISE3.


A
Bacteria;Unclassified;Unclassified;Unclassified;Unclassified;Unclassified;Unclassified
0.00
0.05
0.10
0.15
0.20
Dada
Deblur
Unoise
B
Bacteria;Verrucomicrobia;Unclassified;Unclassified;Unclassified;Unclassified
0.00
0.05
0.10
0.15
0.20
Dada
Deblur
Unoise
C
Bacteria;Verrucomicrobia;Spartobacteria;Unclassified;Unclassified;Unclassified
0.00
0.05
0.10
0.15
0.20
Dada
Deblur
Unoise
D
Bacteria;Acidobacteria;Acidobacteria_Gp1;Gp1;Unclassified;Unclassified
0.00
0.02
0.04
0.06
0.08
0.10
0.12
Dada
Deblur
Unoise
E
Bacteria;Acidobacteria;Acidobacteria_Gp1;Granulicella;Unclassified;Unclassified
0.00
0.02
0.04
0.06
0.08
0.10
0.12
Dada
Deblur
Unoise

Supplemental Figure 7: Top 5 genera driving differences between Deblur and the other two denoising tools in the soil dataset. Boxplots of the relative abundances per sample of five of the classified genera that had relative abundance differences greater than 1% between Deblur and both DADA2 and UNOISE3. Deblur calls more reads that were unclassified at the kingdom and class levels than DADA2 or UNOISE3. A) ASVs only classified at the Bacteria kingdom level. Deblur tends to find higher abundances of these ASVs. B) ASVs only classified at the Verrucomicrobia phylum level. Deblur finds higher abundances of these ASVs. C) ASVs only classified at the Spartobacteria class level. DADA2 and UNOISE3 find more of these ASVs than Deblur. D) ASVs classified at the Gp1 order level of the Acidobacteria_Gp1 class. E) ASVs classified at the Granulicella order level of the Acidobacterta_Gp1 class. Strikingly these two classifications share opposite relationships where Deblur finds more ASVs in the Gp1 order and DADA2 and UNOISE3 find more ASVs in the Granulicella order.

Supplemental Table 1:

Presence of expected organisms across DADA2, UNOISE3, and Deblur in the Human Microbiome Project mock community

Organism	DADA2 Found	Deblur Found	UNOISE3 Found	Precent Abundance
Acinetobacter baumannii ATCC 17978	Yes	Yes	Yes	0.05
Actinomyces odontolyticus ATCC 17982	Yes	Yes	Yes	0.05
Bacillus cereus ATCC 10987	Yes	Yes	Yes	0.05
Bacteroides vulgatus ATCC 8482	Yes	Yes	Yes	0.05
Clostridium beijerinckii ATCC 51743	Yes	Yes	Yes	0.05
Deinococcus radiodurans DSM 20539	Yes	Yes	Yes	0.05
Enterococcus faecalis ATCC 47077	Yes	Yes	Yes	0.05
Escherichia coli ATCC 700926	Yes	Yes	Yes	0.05
Helicobacter pylori ATCC 700392	Yes	Yes	Yes	0.05
Lactobacillus gasseri DSM 20243	Yes	Yes	Yes	0.05
Listeria monocytogenes ATCC BAA-679	Yes	Yes	Yes	0.05
Neisseria meningitidis ATCC BAA-335	Yes	Yes	Yes	0.05
Propionibacterium acnes DSM16379	Yes	Yes	Yes	0.05
Pseudomonas aeruginosa ATCC 47085	Yes	Yes	Yes	0.05
Rhodobacter sphaeroides ATCC 17023	Yes	Yes	Yes	0.05
Streptococcus agalactiae ATCC BAA-611	Yes	Yes	Yes	0.05
Streptococcus mutans ATCC 700610	Yes	Yes	Yes	0.05
Streptococcus pneumoniae ATCC BAA-334	Yes	Yes	Yes	0.05
Staphylococcus aureus ATCC BAA-1718/epidermidis ATCC 12228	Yes	Yes	Yes	0.1

Supplemental Table 2:

Presence of expected organisms across DADA2, UNOISE3, and Deblur in the Extreme mock community

Organism	DADA2 Found	Deblur Found	UNOISE3 Found	Precent Abundance
Bacteroides cellulosilyticus DSM 14838.	Yes	Yes	Yes	4.27E-02
Bacteroides eggerthii	Yes	No	No	4.27E-06
Bacteroides fragilis	Yes	No	No	4.27E-04
Bacteroides massiliensis	Yes	No	No	4.27E-05
Bacteroides ovatus	Yes	Yes	Yes	4.27E-01
Bacteroides thetaiotaomicron	Yes	No	No	4.27E-04
Bacteroides uniformis	Yes	Yes	Yes	4.27E-03
Bacteroides vulgatus	Yes	Yes	Yes	4.27E-01
Barnesiella intestinihominis	Yes	No	No	4.27E-06
Clostridium celatum JCM 1394	Yes	Yes	Yes	4.27E-04
Clostridium cocleatum	Yes	Yes	Yes	4.27E-03
Clostridium methylpentusum DSM 5476	No	No	No	4.27E-06
Clostridium phytofermentans	Yes	No	No	4.27E-06
Clostridium xylanovorans	Yes	Yes	Yes	4.27E-02
Coprococcus comes ATCC 27758	Yes	Yes	Yes	4.27E-03
Eubacterium rectale DSM 17629	Yes	Yes	Yes	4.27E-05
Howardella ureilytica	Yes	No	No	4.27E-06
Parabacteroides distasonis JCM 13400	Yes	Yes	Yes	4.27E-06
Parabacteroides distasonis JCM 13401	Yes	Yes	Yes	4.27E-02
Parabacteroides merdae	Yes	Yes	Yes	4.27E-03
Parabacteroides sp. D13	No	No	No	4.27E-06
Paraprevotella clara YIT 11840	Yes	Yes	Yes	4.27E-05
Prevotella buccalis	No	No	No	4.27E-06
Prevotella copri DSM 18205	Yes	No	No	4.27E-06
Roseburia intestinalis L1-82	Yes	Yes	Yes	4.27E-05
Roseburia inulinivorans DSM 16841	Yes	Yes	Yes	4.27E-04
Ruminococcus gnavus ATCC 29149	Yes	No	No	4.27E-06

Supplemental Table 3:

Presence of expected organisms across DADA2, UNOISE3, and Deblur in the fungal mock community

Organism	DADA2 Found	Deblur Found	UNOISE3 Found	Precent Abundance
Cyberlindnera jadinii	Yes	Yes	Yes	0.0625
Debaryomyces hansenii	Yes	Yes	Yes	0.0625
Diutina catenulata (*Candida catenulata*)	Yes	Yes	Yes	0.0625
Fusarium domesticum	Yes	Yes	Yes	0.0625
Galactomyces geotrichum Type 1	Yes	Yes	Yes	0.0625
Galactomyces geotrichum Type 2	Yes	Yes	Yes	0.0625
Galactomyces geotrichum Type 3	Yes	Yes	Yes	0.0625
Galactomyces geotrichum Type 4	Yes	Yes	Yes	0.0625
Hyphopichia burtonii	Yes	Yes	Yes	0.0625
Kluyveromyces lactis	Yes	Yes	Yes	0.0625
Penicillium allii	No	No	No	0.0625
Penicillium commune	No	No	No	0.0625
Penicillium roqueforti	Yes	Yes	Yes	0.0625
Pichia Kudriavzevii	Yes	Yes	Yes	0.0625
Scopulariopsis fusca	Yes	Yes	Yes	0.0625
Zygosaccharomyces rouxii	Yes	Yes	Yes	0.0625

Supplemental Table 4:

Presence of expected organisms across DADA2, UNOISE3, and Deblur in the Zymomock community

Organism	DADA2 Found	Deblur Found	UNOISE3 Found	Precent Abundance
Bacillus subtilis	Yes	Yes	Yes	0.1571
Enterococcus faecalis	Yes	Yes	Yes	0.1037
Escherichia coli	Yes	Yes	Yes	0.0999
Lactobacillus fermentum	Yes	Yes	Yes	0.1878
Listeria monocytogenes	Yes	Yes	Yes	0.1588
Pseudomonas aeruginosa	Yes	Yes	Yes	0.0462
Salmonella enterica	Yes	Yes	Yes	0.1132
Staphylococcus aureus	Yes	Yes	Yes	0.1331

Supplemental Table 5:

Total ASVs called across DADA2, UNOISE3, and Deblur for all filter stringencies and mock communities

Study	Method	Filter	100% Expected	97% Expected	100% Database	97% Database	Unmatched	Total
HMP	DADA2	High	25	5	0	1	12	43
HMP	Deblur	High	21	0	0	0	0	21
HMP	UNOISE3	High	23	3	0	0	2	28
Extreme	DADA2	High	26	3	33	11	0	73
Extreme	Deblur	High	16	1	9	1	0	27
Extreme	UNOISE3	High	17	2	10	4	0	33
Fungal	DADA2	High	10	1	3	3	2	19
Fungal	Deblur	High	10	0	3	8	6	27
Fungal	UNOISE3	High	10	0	3	13	11	37
Zymomock	DADA2	High	8	3	0	0	0	11
Zymomock	Deblur	High	8	0	0	0	1	9
Zymomock	UNOISE3	High	8	11	0	4	4	27
HMP	DADA2	Med	25	7	0	0	10	42
HMP	Deblur	Med	21	0	0	1	0	22
HMP	UNOISE3	Med	23	3	0	4	4	34
Extreme	DADA2	Med	26	3	34	11	0	74
Extreme	Deblur	Med	16	1	9	1	0	27
Extreme	UNOISE3	Med	17	2	10	6	0	35
Fungal	DADA2	Med	10	1	3	3	2	19
Fungal	Deblur	Med	10	0	3	8	6	27
Fungal	UNOISE3	Med	10	0	3	13	11	37
Zymomock	DADA2	Med	8	3	1	0	0	12
Zymomock	Deblur	Med	8	0	0	1	12	21
Zymomock	UNOISE3	Med	8	15	0	1	19	43
HMP	DADA2	Low	25	7	0	0	9	41
HMP	Deblur	Low	21	0	0	0	0	21
HMP	UNOISE3	Low	23	3	0	4	6	36
Extreme	DADA2	Low	26	3	34	11	0	74
Extreme	Deblur	Low	16	1	9	1	0	27

Extreme	UNOISE3	Low	17	2	10	6	0	35
Fungal	DADA2	Low	10	1	3	3	2	19
Fungal	Deblur	Low	10	0	3	8	6	27
Fungal	UNOISE3	Low	10	0	3	13	11	37
Zymomock	DADA2	Low	8	3	1	0	0	12
Zymomock	Deblur	Low	8	0	0	3	12	23
Zymomock	UNOISE3	Low	8	15	0	1	20	44

Supplemental Table 6:

Total amount of ASVs called by DADA2, UNOISE3, and Deblur in the real soil, BISCUIT, and mouse exercise datasets

Rarefaction	Dataset	UNOISE3	Deblur	DADA2
Before	Soil	12228	8273	23075
After	Soil	11613	8270	16609
Before	BISCUIT	1177	1130	1675
After	BISCUIT	1160	1119	1589
Before	Exercise	1663	1320	2229
After	Exercise	1643	1318	1927