

RESEARCH

Open Access



# The unresolved struggle of 16S rRNA amplicon sequencing: a benchmarking analysis of clustering and denoising methods

Mohamed Fares<sup>1,2†</sup>, Engy K. Tharwat<sup>1†</sup>, Ilse Cleenwerck<sup>3</sup>, Pieter Monsieurs<sup>4</sup>, Rob Van Houdt<sup>5</sup>, Peter Vandamme<sup>6</sup>, Mohamed El-Hadidi<sup>1</sup> and Mohamed Mysara<sup>1\*</sup>

## Abstract

**Background** Although 16S rRNA gene amplicon sequencing has become an indispensable method for microbiome studies, this analysis is not error-free, and remains prone to several biases and errors. Numerous algorithms have been developed to eliminate these errors and consolidate the output into distance-based Operational Taxonomic Units (OTUs) or denoising-based Amplicon Sequence Variants (ASVs). An objective comparison between them has been obscured by various experimental setups and parameters. In the present study, we conducted a comprehensive benchmarking analysis of the error rates, microbial composition, over-merging/over-splitting of reference sequences, and diversity analyses using the most complex mock community, comprising 227 bacterial strains and the Mockrobiota database. Using unified preprocessing steps, we were able to compare DADA2, Deblur, MED, UNOISE3, UPARSE, DGC (Distance-based Greedy Clustering), AN (Average Neighborhood), and Opticlust objectively.

**Results** ASV algorithms—led by DADA2—resulted in having a consistent output, yet suffered from over-splitting, while OTU algorithms—led by UPARSE—achieved clusters with lower errors, yet with more over-merging. Notably, UPARSE and DADA2 showed the closest resemblance to the intended microbial community, especially when considering measures for alpha and beta diversity.

**Conclusion** Our unbiased comparative evaluation examined the performance of eight algorithms dedicated to the analysis of 16S rRNA amplicon sequences with a wide range of mock datasets. Our analysis shed light on the pros and cons of each algorithm and the accuracy of the produced OTUs or ASVs. The utilization of the most complex mock community and the benchmarking comparison presented here offer a framework for the comparison between OTU/ASV algorithms and an objective method for the assessment of new tools and algorithms.

**Keywords** Operational taxonomical units (OTUs), Amplicon sequence variants (ASVs), 16S rRNA amplicon sequencing, Denoising.

<sup>†</sup>Mohamed Fares and Engy K. Tharwat contributed equally to this work.

\*Correspondence:  
Mohamed Mysara  
MMaysara@nu.edu.eg

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## Background

16S rRNA amplicon sequencing is a powerful tool to infer the microbial composition of a given sample [1, 2]. This genomic marker is sequenced with appropriate depth to identify the microbial members and their relative abundance. However, this approach is vulnerable to biases and technical errors introduced during several steps of this protocol, such as contaminant sequences, PCR point errors, chimeric artificial sequences, and sequencing errors [3–6]. Sequencing errors, in particular, are platform-dependent, with Illumina sequencing primarily exhibiting nucleotide substitutions rather than indel errors seen with other platforms. These substitutions often stem from high correlation in signal intensities, which are also influenced by signal dependency between cycles or the presence of the GC-rich motifs [4]. Consequently, the presence of erroneous reads and artifacts affects the observed diversity drastically, thereby posing a significant challenge to the identification of biological reads that truly represent the members of the microbial community members [7, 8].

Typically, clustering reads based on their identity (usually 97%) into a single taxon has been used to overcome this sequencing noise, which is commonly referred to as an operational taxonomical unit (OTU). This approach assumes that these variants originate from one genuine biological sequence that has been affected by errors introduced during the sequencing process [9, 10]. Among the clustering-based methods, UPARSE and VSEARCH-DGC implement a greedy clustering algorithm to construct the OTU structure [11, 12]. On the other hand, mothur software calculates a distance matrix that is clustered with a fixed similarity cutoff using either nearest, furthest, or average neighbor methods [10]. A recent algorithm in mothur, named Opticlust, assembles the clusters iteratively, evaluates their quality through the Matthews correlation coefficient, and consequently merges, relocates, or assigns as a novel cluster [13].

Despite the advancements in those clustering algorithms, concerns remain about the limitations of utilizing a rigid clustering cutoff [14]. Previous work demonstrated that the clustering cutoff should be region-specific and taxa-dependent, as applying a more stringent cutoff might encompass distinct taxonomical variations, while a more relaxed cutoff could fail to capture meaningful biological insights [15, 16]. Nonetheless, applying such dynamic cutoffs is dependent on the extent to which a microbial niche is studied, leaving the 3% cutoff as the default for most clustering algorithms [17–25].

More recent methods have emerged to address this problem by producing Amplicon Sequence Variants (ASVs). These techniques rely on different statistical models to discriminate real sequences from spurious ones. Tools such as DADA2, MED, UNOISE3,

and Deblur adopt this approach. DADA2 implements an iterative process of error estimation and partitioning sequences based on the model implemented [26, 27]. Minimum Entropy Decomposition (MED) relies on a similar iterative premise by detecting sequence-position entropies not likely to be explained by errors [28, 29]. Deblur employs a pre-calculated statistical error profile to estimate the position's likelihood of being erroneous and corrects it accordingly [30]. UNOISE3 compares the abundance of reads to similar sequences and then collapses identical reads into error-free and erroneous categories, utilizing a probabilistic model to assess insertion and substitution probabilities for denoising. Such denoising approaches are presumed to improve the taxonomical resolution by differentiating biological differences on a single-nucleotide level. An apparent advantage is the consistency of ASVs as sequence labels that can be used across studies without the need for re-clustering, in contrast with OTU approaches [16]. Nevertheless, this method introduces other problems, for instance generating several ASVs for non-identical 16S rRNA gene copies within the same strain. More recent methods emerged to address this problem by producing Amplicon Sequence Variants (ASVs), Exact Sequence Variants (ESVs), or zero radius OTUs (zOTUs), which are all referring to the same concept. These techniques rely on different statistical models to discriminate real sequences from spurious ones. Further details on the various OTUs/ASVs approaches are provided in Additional File 1.

Several studies have compared their developed clustering and denoising pipelines to existing alternatives; however, few independent studies have comprehensively compared clustering methods without focusing solely on algorithmic concepts or including comparisons at the ASV algorithm level [31–38]. Consequently, varying filtering criteria and chimera removal methods within these pipelines obscured a clear understanding of the performance of the various clustering/denoising approaches [33]. Another challenge comes with the selection of the appropriate dataset for the comparative analysis, as argued by Bokulich et al. [39]. Although numerous microbial community samples exist, the absence of the ground truth impedes our ability to properly evaluate the clustering/denoising approaches [27, 35, 40, 41]. While mock community data is an appropriate alternative, available mock datasets do not capture the complexity naturally found in diverse biological environments [see Additional file 2]. Thus, there is a need for an objective comparative analysis utilizing a more complex mock sample able to independently evaluate the strengths and limitations of the clustering/denoising approaches.

In this work, we aim to subject clustering and denoising approaches to an unbiased and challenging head-to-head comparison and to highlight the strengths and limitations

of each approach [see Additional file 1]. In particular, we will explore their differences by predicting the composition of the different mock communities, their error rate, merging/splitting behavior, diversity analyses, and run-time analysis. For this purpose, we used 16S rRNA amplicon sequencing data generated from the most complex microbial mock community to date, consisting of 227 bacterial strains from 197 different species [42] covering V3–V4 region. In addition, we included a plethora of publicly available mocks that cover V4 region alone to enrich our comparative analysis and to further illustrate the differences between the above-mentioned algorithms.

## Methods

### Mock data

Two main data sources were used in this study. The first dataset, HC227\_V3V4, was generated by amplifying the HC227 mock community with primers targeting the V3–V4 variable region of the 16S rRNA gene (5'-CCTACGGGNGGCWGCAG-3' and 5'-GACTACH-VGGGTATCTAATC-3' as forward-reverse primer pair). Sequences were obtained on an Illumina MiSeq4000 platform in a 2×300 bp paired-end run (Illumina Inc., San Diego, CA, USA). The HC227 mock community consists of genomic DNA from 227 bacterial strains belonging to 197 different species [42]. The second dataset consisted of thirteen 16S rRNA gene amplicon datasets [see Additional file 3] collected from the Mockrobiota database [39]. Paired-end Illumina-MiSeq mock data sets were collected along with their expected reference sequences and taxonomic reference composition. The mocks covered a wide spectrum of input diversity ranging from 15 to 59 bacterial species; we selected the mock samples covering the V4 region of the 16S rRNA gene to reduce discrepancies. The 16S rRNA gene-targeted regions of the reference species within HC227\_V3V4 and Mockrobiota [39] mock communities were dereplicated using the *unique.seqs* command in *mothur* (v.1.43.0) [10], which yielded distinct reference-variants referred to as *ASV-ref*. Subsequently, these unique sequences for both datasets were clustered into OTUs, which we referred to as *OTU-ref*.

### Data preprocessing

Sequence quality was checked using FastQC (v.0.11.9, Babraham Bioinformatics); primer sequences were stripped using the *cutPrimers* (v.2.0) tool [43]. Paired-end reads were merged using the *fastq\_mergepairs* command in USEARCH (v 11.0.667) [44], and length trimming was achieved using PRINSEQ tool (v 0.2.4) and FIGARO [45, 46]. Misoriented reads were aligned to the SILVA database (Release 132) [47, 48], and incorrectly oriented (i.e., flipped) reads were filtered out using the *screen.seqs* command in *mothur* (v.1.43.0) [10]. Further quality filtration

was performed using the USEARCH (v11.0667) *fastq\_filter* command to discard all reads possessing ambiguous characters as well as optimize the maximum error rate *fastq\_maxee\_rate*=0.01 [44]. Unlike the rest of the tools, DADA2 read merging occurs towards the end of the filtering steps. Thus, to detangle the effect of merging from the denoising/clustering effect (our main objective), reads were analyzed in two scenarios, one using only the forward reads as single-end (SE) reads and a second with merged forward and reverse paired-end reads (PE). This made it possible to assess whether the observed performance is different between PE and SE reads. Mock samples were subsampled to 30,000 reads per sample using the *mothur sub.sample* command to have a reasonable level of errors/artifacts.

### Methods included in the comparison

We compared the performance of four ASV denoising approaches, i.e., DADA2 (v 1.16) [27], Deblur (v 1.1.0) [30], MED (v 2.0) [29], and UNOISE3 (v11.0.667) [49], and four clustering methods, i.e., UPARSE (v11.0.667) [11], average neighborhood (AN) (v 1.43.0) [10], Opticlust (v 1.43.0) [13], and VSEARCH (v 1.43.0) [12]. All the parameters set for each algorithm can be found in Additional file 4.

### Sequence error correction and OTU generation

For DADA2, the error model was trained using the *learnErrors* command, and sequence variant inference was done using the core command *dada* (setting *OMEGA\_C* parameter was set to 0). As DADA2 is highly dependent on its own preprocessing, another set of the results is done using the default parameters together with the unified parameters. In Deblur, filtering was performed using the default Deblur positive mode, and biom tables were built using chimera retained denoised files. For MED, the default pipeline command *decompose* was applied with the minimum substantive abundance parameter set to one, *relocate-outliers* and *skip-removing-outliers* options were turned on to disable the removal of outlier and low abundance reads.

For UPARSE and UNOISE3, the *fastx\_uniques* command was used to identify unique sequences along with their abundances using default parameter settings. In UPARSE, the reads were clustered using the *cluster\_otus* command. In UNOISE3, reads were denoised using the *unoise3* command. For both, the *minsize* parameter was set to one, chimeric reads were retained, and the OTU table was generated using the *otutab* command.

### Clustering workflow for mothur algorithms

The Average neighborhood, VSEARCH (calculated using distance-based greedy clustering; DGC), and Opticlust methods were used within the *mothur* pipeline. For all

three methods, input FASTA formatted files were dereplicated using the *unique.seqs* command and unique read counts were reported using the *count.seqs* command. Preliminary clustering was performed using the *pre.cluster* command and the pairwise distance matrix was calculated using the *dist.seqs* command. The three clustering approaches were applied separately using the *cluster* command with a 0.03 distance cutoff.

Finally, shared files were generated using *make.shared* for each method and OTU FASTA files were generated using the *get.oturep* with the method parameter set to *abundance*. To standardize the comparison, chimera removal method was unified using the *seq.error* command against the mock's reference sequences to avoid discrepancies between chimera-removal algorithms. Additionally, singleton reads, i.e., reads with only one occurrence, were retained for all methods (Fig. 1).

### Analysis and output comparison

We applied several criteria to evaluate the results obtained from the different algorithms. Considering that each approach received the same input (after the unified pre-processing), the generated OTUs/ASVs could be put head-to-head against several criteria, namely microbial composition, error rate, and OTUs/ASVs boundary definition. Also, Shannon and Observed Feature indexing for alpha diversity were calculated using the *summary.single* command in mothur. Beta diversity was calculated using

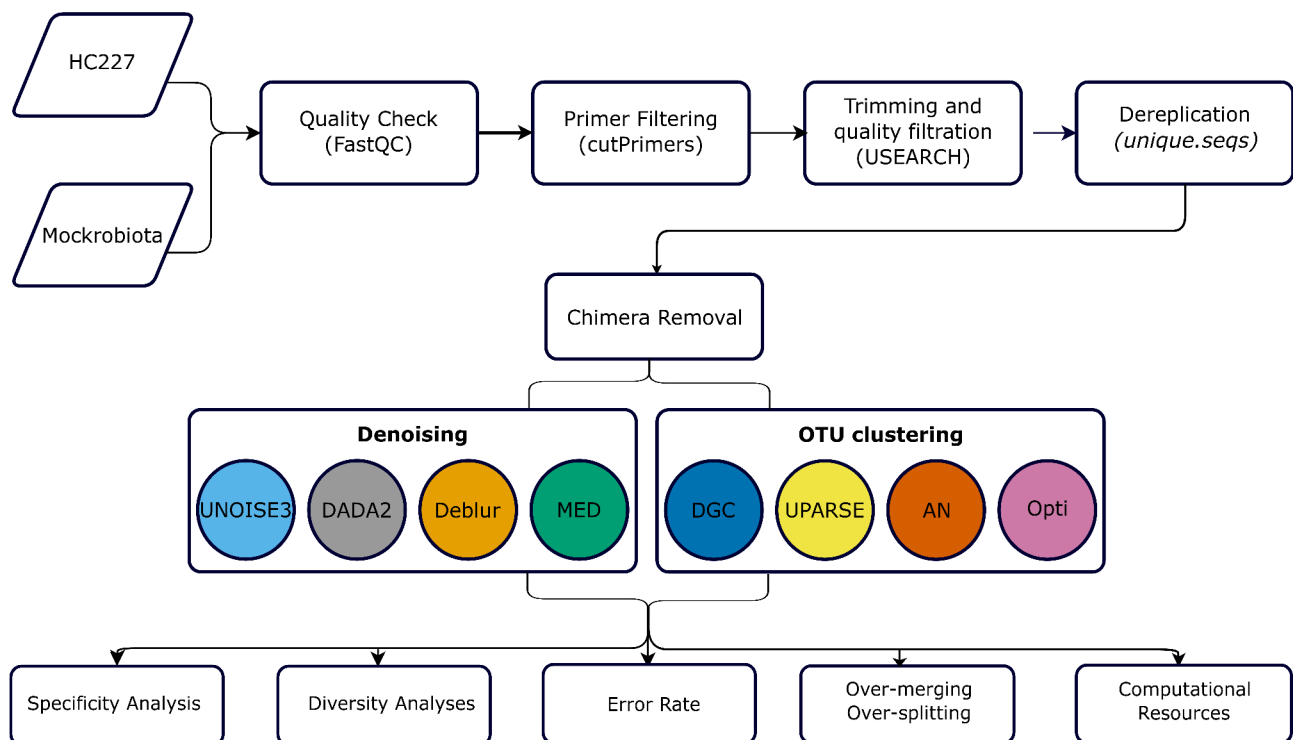
commands *dist.shared* and *pcoa*, respectively. To further assess sequencing depth, rarefaction curves were generated, and alluvial plots were created to visually represent the diversity outputs of each algorithm compared to both ASV-ref and OTU-ref ground truth versions.

### Specificity analysis

OTU tables generated by different approaches were parsed into the mothur count file format and representative OTU sequences were generated. Subsequently, these OTUs/ASVs were categorized into four groups according to reference sequence identity: exact match (100% sequence identity with the reference sequence), mismatch (>97% sequence identity with the reference sequence), contamination (<97% sequence identity with the reference sequence and >97% sequence identity with a SILVA database entry), and others (all not classified above). The sequence identity was calculated using standalone BLAST [v 2.10.0; 50] with query coverage set to 100%.

### Error rate calculation

The PCR amplification and sequencing process inherently poses the possibility of introducing several biases, such as PCR single-base mistakes, PCR chimeras, and sequencing errors. According to recent work [51], it is essential to carefully identify and eliminate these mistakes as part of the analytical process. The research



**Fig. 1** Overall workflow of benchmarking steps. The process starts from data pre-processing, followed by dereplication, denoising or clustering and chimera removal, ending with OTUs & ASVs tables and evaluation comparisons performed on them

results may be skewed if these biases are not adequately addressed, which could result in a large, inaccurate estimation of the diversity of the microbial community. For each algorithm, chimeric reads were determined using the *mothur* command *seq.error* command and removed before the error rate of the OTUs/ASVs was calculated for each mock. The error rates of the OTUs/ASVs were assessed against reference sequences in terms of mismatches, insertions, and deletions and calculated by comparing the number of mismatched positions to the total number of bases for each mock.

### Merging/splitting analysis

Furthermore, the over-merging rates (where more than one reference sequence was merged into the same OTU/ASV) and over-splitting frequencies (where the same reference sequence was split into multiple OTUs/ASVs) were evaluated. For this purpose, the OTUs/ASVs were mapped against the unique and clustered reference sequences (i.e., *ASV-ref* and *OTU-ref*, respectively) using standalone BLASTn (v 2.10.0) [50], setting query coverage to 100% and percent identity to 97%. Reference sequences that were found absent, albeit being present prior to the denoising/clustering step, were considered over-merged. Moreover, if the same reference sequence was mapped to more than a single OTU/ASV, it was considered over-split. Furthermore, a head-to-head comparison was done comparing the abundance of each bacterial strain in the HC227\_V3V4 mock community against the number of sequence reads aligned to it using compatible ASV/OTU reference versions.

### Diversity analysis

To compare the diversity of observed OTUs/ASVs across different approaches, we generated rarefaction curves using the *rarefaction.single* command in *mothur*. In terms of diversity analysis, we calculated alpha diversity using Shannon indexing [52], which considers both the abundance and evenness of species present and Observed Feature (Sobs Calculator) indexing [52], which only considers evenness. The *summary.single* command in *mothur* was utilized for calculating both indexing metrics. For beta diversity, we calculated distances between the various approaches and the theoretical (i.e., expected) *ASV-ref* and *OTU-ref* abundance using *mothur summary.single*, and *dist.shared* commands then we compared generated coordinates by PCoA plotting using *pcoa* command. We accommodated Jclass and Euclidean distance metrics [53, 54] that focus on the presence or absence of ASVs rather than their abundance and Canberra similarity calculation [55, 56] that consider both presence/absence and abundance. Moreover, we conducted a direct comparison of beta diversity distances between

the algorithms against both our reference versions (i.e., *OTU-ref* and *ASV-ref*) as benchmarks.

### Computational cost analysis and clustering parameter effect

To evaluate the runtime and memory consumption of each clustering/denoising algorithm, we conducted tests using subsampled, preprocessed reads from the HC227\_V3V4 mock community. Using different sets of 5000 to 20,000 reads per sample utilizing the *sub.sample* command in *mothur* with varying size parameters. Each algorithm was executed on a PC equipped with 64 GB of RAM and 12 threads, with multithreading enabled across all algorithms. Moreover, we tested cutoffs ranging between 0.01 and 0.05 for clustering-based approaches, with exception of UPARSE, as it was not possible to change the hardcoded cutoff.

### Statistical analysis

We statistically evaluated the error rate, specificity, and merging/splitting differences across various OTU/ASV algorithms. Normality of the data was assessed using the Shapiro-Wilk test. For normally distributed data, we applied ANOVA followed by Tukey's HSD test, whereas for non-normally distributed data, we used the Kruskal-Wallis test followed by the Wilcoxon rank-sum test. A significance level of 0.05 was applied to corrected p-values. All statistical analysis and graphs were constructed using R (v.4.2.2), patchwork [57], ggplot2 [58], ggalluvial [59], and tidyr [60] packages. Codes and scripts for analysis and visualization are available at GitHub page: <https://github.com/MOFares-Bioinf/BACDAS>.

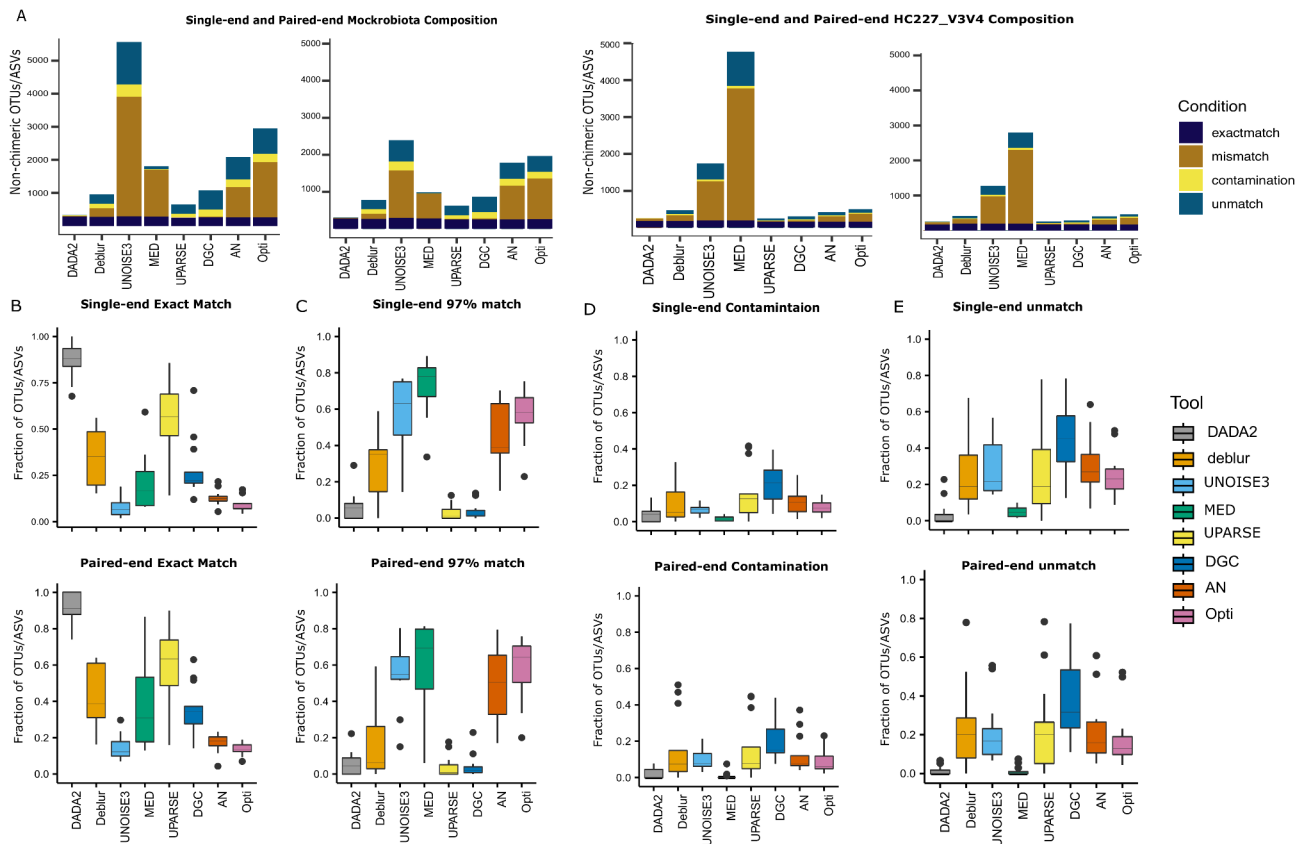
## Results

### DADA2 and UPARSE provide the most accurate estimate of microbial composition

The performance of the various denoising/clustering approaches in terms of specificity analysis was assessed using the HC227\_V3V4 and Mockrobiota mocks in the SE and PE scenarios. For the HC227\_V3V4 dataset, DADA2 (whether using the unified or the default parameters) and UPARSE had the highest count of exact matches, followed by DGC, AN, Opticlust, and Deblur. This was consistent for both single-end and paired-end reads and in agreement with the results from the collection of 13 mocks within the Mockrobiota community, except for DGC, AN, and Opticlust (Fig. 2A-B; Additional file 5; Supplementary Fig. 1).

DADA2 had 6% of its ASVs categorized as mismatches for the Mockrobiota community, which was higher compared to the 1% and 2% observed for UPARSE and DGC, respectively (Fig. 2C). Nonetheless, MED and UNOISE3 had the highest percentage of mismatches for the HC227\_V3V4 mock community, with percentages of





**Fig. 2** Illustrating Performance of the various algorithms in respect to the microbial composition. Stacked bar plots of OTUs/ASVs output composition for single-end and paired-end reads from Mockrobiota and HC227\_V3V4, respectively (A). Average box plots showing the fraction of exact matches (B), 97% matches (C), contaminants (D) and unmatched (E) for single-end and paired-end reads for the Mockrobiota mock community

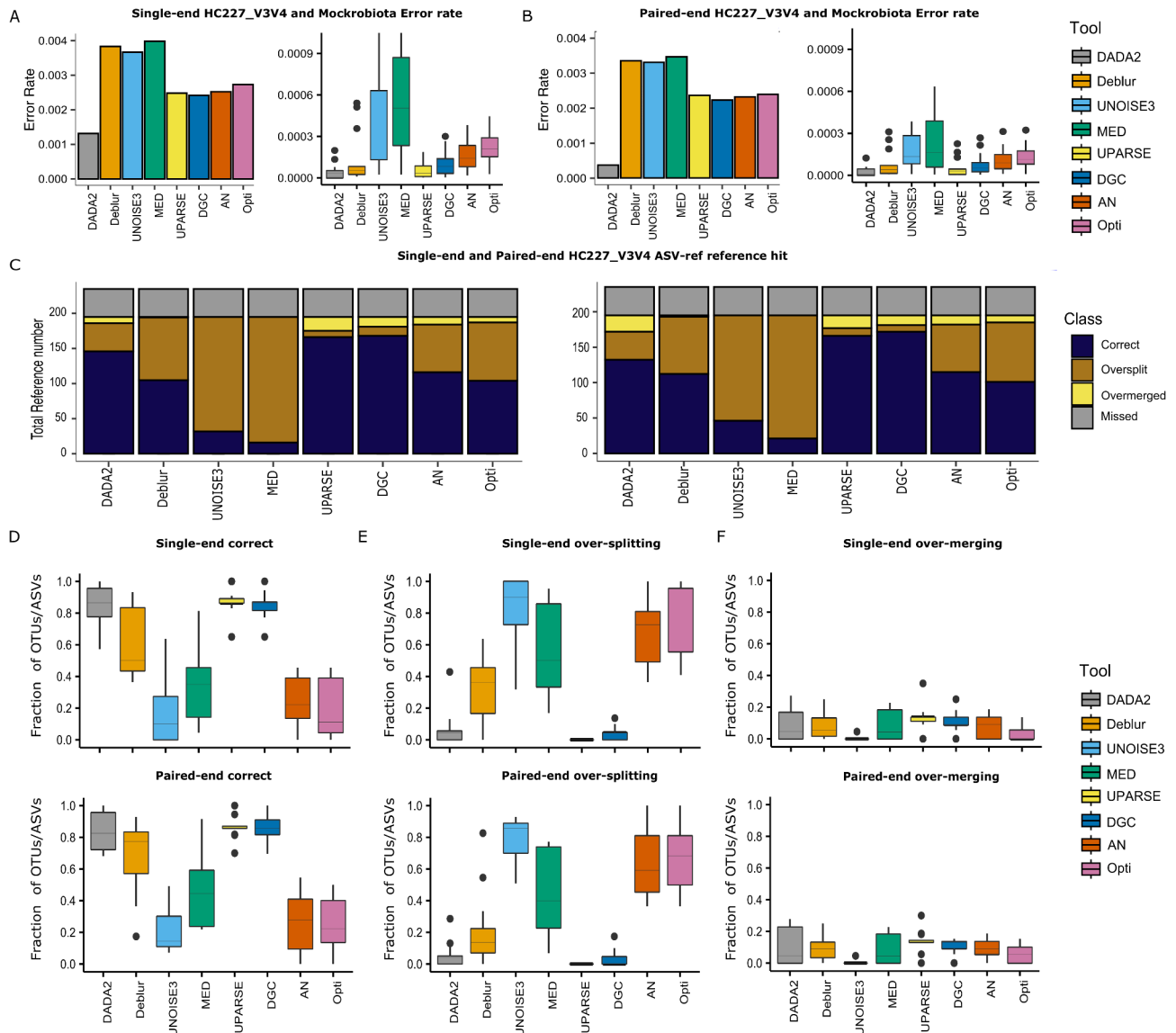
75% and 61%, respectively (Fig. 2A). This outcome was consistent with the findings observed within the Mockrobiota dataset (Fig. 2C; Additional file 5). Statistical validations across the eight algorithms for each category in the Mockrobiota dataset are provided in Additional file 5. When looking into DADA2 using the default parameters, mismatch level was higher for both Mockrobiota and HC227\_V3V4 mock communities relative to DADA2 unified parameters (Supplementary Fig. 1–A & C).

When examining the OTUs/ASVs identified as potential contaminants, approximately 3% were detected by DADA2 (in unified and default parameters), MED, and UNOISE3 in the Mockrobiota mock community. In contrast, UPARSE and DGC identified fewer contaminants, with less than 10%, in the Mockrobiota community compared to the HC227\_V3V4 mock community. The results were consistent for both single-end and paired-end reads (Fig. 2D; Supplementary Fig. 1 - C). Finally, all approaches achieved approximately the same count of unmatched reads, around  $20 \pm 4\%$ , except DADA2, that performed better with only 5% for HC227\_V3V4 (Fig. 2A). Yet, this was not consistent for DGC within the Mockrobiota dataset, where it surged above 40%, and MED achieved below 5% unclassified OTUs (Fig. 2E).

### DADA2 and UPARSE showed the lowest error rate yet still with some over-merging in the ASV-ref

Next, the error rate of the OTUs/ASVs from the various clustering/denoising approaches was assessed (Fig. 3: A-B). DADA2, UPARSE, and DGC demonstrated the lowest overall value, while MED and UNOISE3 exhibited the highest error rate for both the HC227\_V3V4 and Mockrobiota community datasets (Fig. 3A-B). Interestingly, Deblur's performance was inconsistent between the two datasets, with a high error rate for the complex HC227\_V3V4 mock and a much lower error rate for the Mockrobiota mocks. For DADA2, the error rate of both datasets increased when utilizing the default parameters instead of the unified parameters (Supplementary Fig. 1-B). The AN and Opticlust algorithms showed approximately the same error rate for both datasets. Statistical validation was performed for Mockrobiota dataset, ensuring the robustness of the results.

In addition, by comparing the OTUs/ASVs against the mock reference sequences, it was possible to assess whether each reference variant was mistakenly grouped with others into the same OTU/ASV (over-merged) or split over multiple OTUs/ASVs (over-split; Fig. 3: C-F). Regarding the number (and fraction) of correctly



**Fig. 3** Illustration of various algorithms performance for evaluating error rate and merging/splitting based **ASV-ref**. **A, B**) Bar plots demonstrate the error rate for HC227\_V3V4, and Box plots demonstrate the error rate for Mockrobiota community across the algorithms for single-end and paired-end reads. **C, D**) Stacked bar plots demonstrates single-end and paired-end for total number of references in the ASV-ref for Mockrobiota and HC227\_V3V4 respectively. **E, F**) Average box plots demonstrate correct percentage for single-end and paired-end Mockrobiota community, **G, H**) Average box plots demonstrate over-splitting percentage for single-end and paired-end reads for Mockrobiota community, **I, J**) Average box plots demonstrate Over-merging percentage for single-end and paired-end reads for Mockrobiota community

assigned OTUs/ASVs, UPPARSE and DGC showed the highest performance, followed by DADA2 for both communities (Fig. 3: C, D). On the other hand, MED, UNOISE3, and Deblur identified the lowest number of correctly assigned ASVs for both datasets, while Opticlust and AN showed a slightly better performance for HC227\_V3V4 and Mockrobiota Community. Similar results were reported for over-splitting, with UPPARSE showing a marginal to negligible over-splitting followed by DGC and DADA2 for both datasets (Fig. 3: C, E).

Most approaches suffered from varying degrees of over-merging, except for Opticlust and UNOISE3, which

showed less to approximately no over-merging in both datasets. Yet, the worst performance regarding the over-merging was achieved by MED followed by DADA2 which showed a noticeably higher level of over-merging when relying on paired-end than on single-end reads (Fig. 3: C, F). The performance of DADA2 deteriorated when utilizing default parameters, showing higher levels of over-merging for Mockrobiota mock community and higher levels of over-splitting with a slight reduction in correctly identified reads in HC227\_V3V4 mock community (Supplementary Fig. 1:E). Deblur showed a slightly better performance for the Mockrobiota community

dataset compared to the HC227\_V3V4 dataset, which was consistent with its error rate results for the Mockrobiota community.

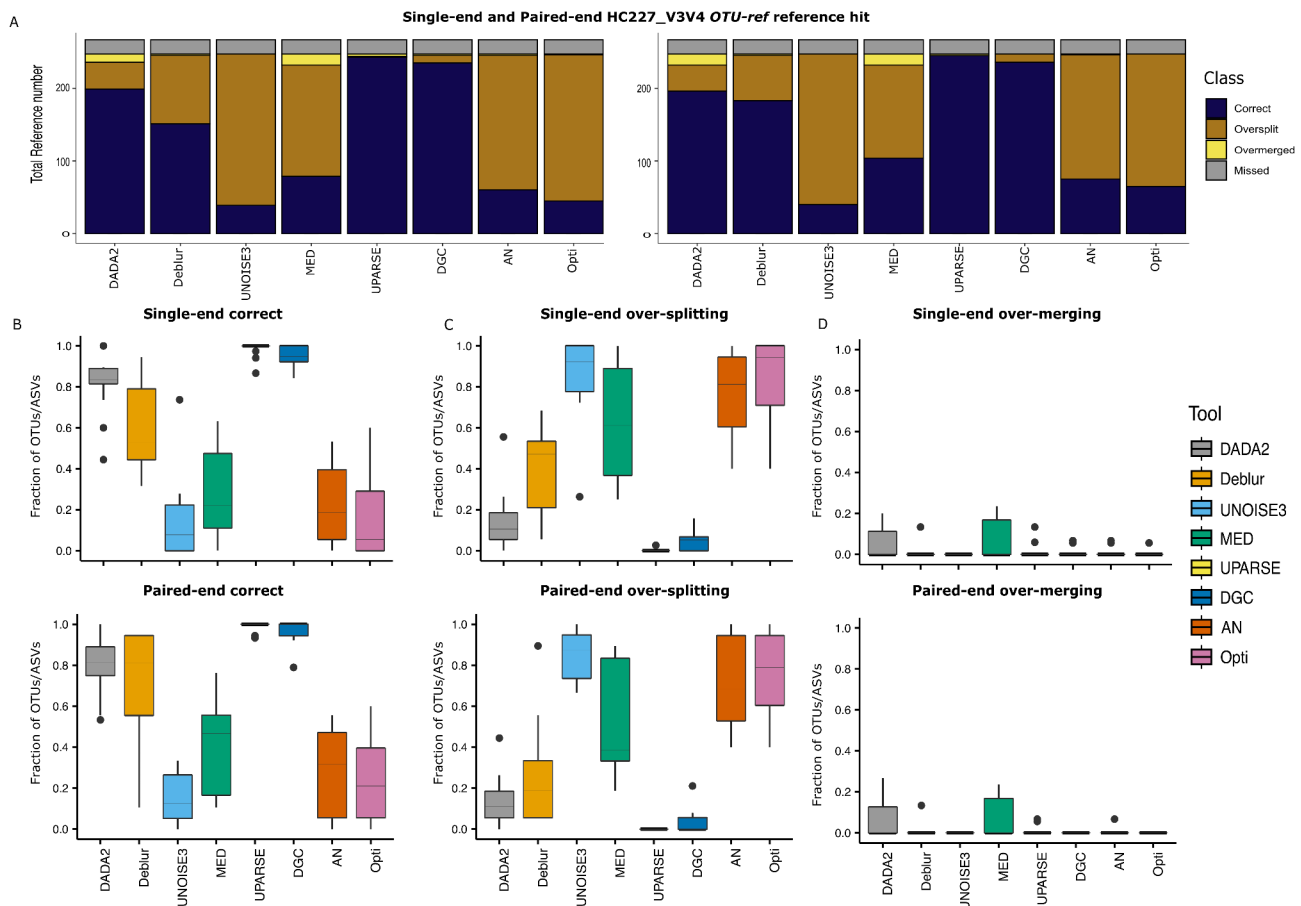
#### UPARSE, DGC, and DADA2 had the highest accuracy with least fraction of over-splitting and over-merging in the *OTU\_ref*

When comparing the resulting OTUs/ASVs to the clustered mock references, UPARSE, DGC, and DADA2 exhibited the highest percentage of correctly assigned reads and the lowest over-split percentage for both the Mockrobiota and HC227\_V3V4 mock communities (Fig. 4 and Additional file 5 for the statistical analysis results). For AN, Opticlust, and UNOISE3 algorithms, only around 20% of the reference sequences were correctly matched with one unique OTU/ASV in the HC227\_V3V4 dataset, which suggests that similar reads originating from the same reference were clustered separately (Fig. 4A). Yet, most reference sequences were split

over exceeding a single OTU/ASV per reference for these algorithms.

Furthermore, Deblur showed an average performance with 35–40% of the reference sequences correctly matched with a single OTU/ASV percentage. Yet, this percentage was significantly increased for the Mockrobiota dataset (Fig. 4B). The results were mostly identical for single-end and paired-end reads [see Additional file 4]. The UNOISE3, Opticlust, and AN algorithm, with their large number of OTUs/ASVs, showed a high percentage of over-splitting reads compared to other approaches for the Mockrobiota community (Fig. 4C). Yet, UNOISE3 and MED showed the highest percentage of over-splitting OTUs/ASVs among the tested approaches for the HC227\_V3V4 dataset (Fig. 4A).

Notably, 17% of HC227\_V3V4 and 40% of Mockrobiota reference sequences were missed by all tested approaches. Although over-merged ASV-references were present in all clustering-based algorithms, this was effectively resolved when using *OTU\_ref* as originally intended



**Fig. 4** Illustration of various algorithms performance using clustering method for evaluating over-merging/over-splitting based *OTU\_ref*. **(A)** stacked bar plots represent total number of reference sequences for single-end and paired-end conditions for Mockrobiota and HC227\_V3V4 datasets respectively. **(B)** Average box plots representing the correct percentage for single-end and paired-end reads for Mockrobiota community **(C)** Average box plots representing the over-splitting percentage for single-end and paired-end reads for Mockrobiota community, **(D)** Average box plots representing the over-merging percentage for single-end and paired-end reads for Mockrobiota community



for these algorithms [see Additional file 6]. In contrast, for ASV algorithms—MED in particular—this was not the case as shown above.

After evaluating different clustering cutoffs on clustering-based algorithms compared to *OTU-ref*, cutoff 0.03 was the optimal one, yielding the highest number of correctly assigned species for all clustering approaches. Moreover, a notable increase in over-merging and decrease in over-splitting percentage was observed as less strict cutoff was applied [see Additional file 7].

#### **DADA2 and Deblur showed the closest resemblance to the mock composition**

We analyzed the composition of the HC227\_V3V4 mock community and calculated the theoretical reference composition (*OTU-ref* and *ASV-ref* compatible with the OTU and ASV approaches, respectively). Regarding Shannon's alpha diversity index, all approaches showed a very close resemblance to the *ASV-ref* or *OTU-ref*, with exception of MED and UNOISE3, which was consistent with their sub-optimal performance in respect to the reported error rates and the inflation of over-split ASVs for the same references (as shown before). For the observed feature's richness index, DADA2 and deblur (unlike MED and UNOISE3) showed the closest microbial diversity to our input. These results were consistent with sequence mapping analysis against *ASV-ref* for the denoising algorithms [see Additional file 6]. For the clustering approaches, UPARSE showed the closest resemblance to the mock composition (*OTU-ref*), followed by DGC, AN, and Opticlust algorithms, in line with the observed feature and shannon matrices results (Fig. 5).

For beta diversity, both membership-only distances (e.g., Jclass and Euclidean matrices) and membership with abundance-based metrics (e.g., Canberra) showed that DADA2 and Deblur had the closest resemblance to the *ASV-ref* input (Supplementary Fig. 2), while UNOISE3 and MED exhibited less similarity and greater variance in comparison to the *ASV-ref* input. This was consistent for single-end and paired-end approaches (Fig. 5C). When evaluating the clustering algorithms against the *OTU-ref* ground truth, UPARSE exhibited the closest resemblance to the *OTU-ref* ground truth, with a high number of shared OTUs. This performance was followed by DGC, AN, and Opticlust. Notably, the results were consistent across both single-end and paired-end sequencing approaches (Fig. 5D).

We also compared *ASV-ref* input against denoisers and *OTU-ref* input against clustering algorithms via alluvial plots. DADA2 seems to preserve notable portion of the diversity found in the *ASV-ref* input compared to the other denoisers, as shown in Supplementary Fig. 3, which is consistent with the alpha diversity results (Fig. 5A).

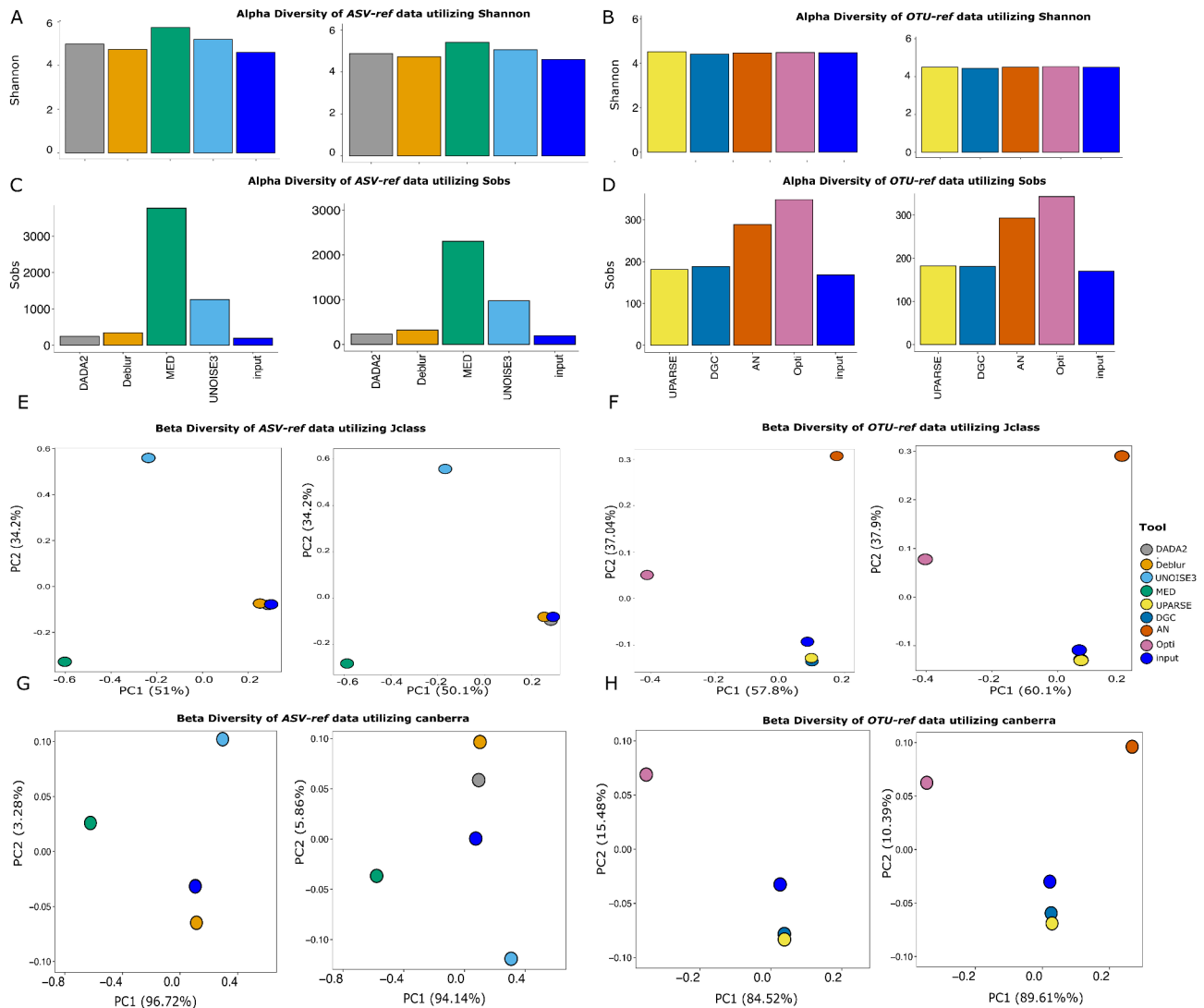
For clustering algorithms, UPARSE seems to maintain a large amount of diversity of *OTU-ref* references and seems to balance between over-merging and over-splitting with discrepancies from the reference OTUs. Opticlust algorithm also maintains a large number of *OTU-ref*, yet is prone to over-splitting, as shown in Supplementary Figs. 2 & 3, which is also consistent with over-splitting results in Fig. 4A. The behavior of both denoising and clustering algorithms was consistent between single-end and paired-end scenarios [see Additional file 8].

#### **Computational resources and parameter effect exploration**

Analysis of execution time, considering both CPU processing and input/output operations, showed that all algorithms performed similarly, with execution times gradually increasing as the number of input sequences grew. Yet, MED and Deblur required execution times ranging from 2 min to 120 min compared to the other algorithms, which required between 2 s and 4 min [see Additional file 9]. Furthermore, the RAM amount required for all clustering/denoising algorithms was between 5 and 11 GB of RAM, except for MED, which required up to 40 GB of RAM for a sample containing 20,000 reads. Also, a gradual increase in memory consumption was noted with the increase in the number of input sequences.

#### **Discussion**

Defining species boundaries using 16S rRNA amplicon sequencing presents a significant challenge [15]. Analysis mainly adopts one of two approaches: either clustering similar sequences within a certain distance threshold (referred to as OTU-based approaches, such as DGC, AN, Opticlust, and UPARSE) or denoising erroneous reads (referred to as ASV-based approaches, such as DADA2, Deblur, MED, and UNOISE3). Biases and variability in different preprocessing steps make it challenging to compare such algorithms [39]. Additionally, mock community samples are typically used for this purpose, yet the available mock communities do not capture the true microbial complexity, as argued before [39]. To accommodate this, we sequenced 16S rRNA amplicons of the most complex mock community previously described in Goussarov et al. [42] and used it together with a collection of thirteen previously available mocks for a comprehensive comparative analysis. We used unified preprocessing and filtering steps to allow benchmarking the clustering/denoising algorithms objectively. As DADA2 error profiling might be sensitive to its own preprocessing, we also provided the results using the default parameters. We adopted both single-end and paired-end read scenarios in our comparison to ensure an accurate evaluation, particularly for Deblur designed for single-end reads.



**Fig. 5** illustration of Microbial Composition, alpha and beta Diversity for HC227\_V3V4 dataset. **A, B**) Bar plots for Shannon and Observed Feature indexing calculation for Denoising and clustering approaches compared to ASV-ref and OTU-ref theoretical input for single-end and paired-end conditions respectively. **C, D**) scatter plots for comparing Denoising and clustering approaches to both ASV-ref and OTU-ref theoretical input respectively using Jclass and Canberra calculation

Although all algorithms were allowed to identify error-free OTUs/ASVs (exactly matching the intended reference species within the mocks), they varied drastically in their ability to handle erroneous reads. DADA2, UPPARSE, and DGC achieved the highest performance, with the lowest number of erroneous OTUs/ASVs and consequently the lowest number of overall OTUs/ASVs. MED and UNOISE3 had the highest number of mismatches, which exceeded the other algorithms by several folds. Interestingly, when using the default parameters, DADA2's performance deteriorated—producing fewer ASVs that exactly matched the Mockrobiota mock community while generating additional ASVs with mismatches in both the Mockrobiota and HC227\_V3V4 mock communities. This was in line with

previous whole-pipeline comparative studies, where DADA2's pipeline was more capable of identifying ASVs exactly matching the reference sequence than Deblur and UNOISE3, yet at the expense of increased over-splitting [37, 38]. Yet, in our analysis, we were able to pinpoint this phenomenon to the influence of the denoising algorithms, which was not possible previously due to the confounding effect of applying different filtering and preprocessing parameters for each approach's pipeline. The lower percentage of contamination and unmatched sequences in denoising compared to clustering algorithms might suggest strict handling/filtering of relatively low abundance sequences, such as contamination and chimeras, by ASV approaches as also noted by Reitmeier et al., 2021 [61].

The algorithm's efficiency in eliminating OTUs/ASVs with artifacts can also be assessed by the error rate. For instance, DADA2, UPARSE, and Deblur showed the lowest error rates in the produced OTUs/ASVs, which was consistent with their accurate assessment of the mocks' microbial community. Yet, the error rate of DADA2 increased for both mock communities when using the default parameters compared to the unified parameters. Also, accurately correcting errors requires extensive memory usage and processing time for ASV approaches, which might be attributed to the error correction and modeling steps involved. Consequently, UNOISE3 and MED had the highest error rates, reflecting their lower performance in assessing the microbial community. Although this difference in error rate between DADA2 and MED/UNOISE3 seems minor, it had a profound effect on the biological interpretation. For instance, findings of MED and UNOISE3 diverged significantly from those produced by DADA2 and UPARSE. The former two identified the least number of correctly assigned bacterial species in addition to the largest deviation from the input data, as also reflected by both observed feature and shannon indices. The same holds for beta diversity utilizing three different distance matrices.

The other clustering-based algorithms, DGC, AN, and Opticlust performed moderately, with noticeable differences between the mean error rates of the tested mocks. Furthermore, none of the tested algorithms allowed retrieving the complete number of expected mock reference sequences, as up to 17% for HC227\_V3V4 and 40% for Mockrobiota were missed by all tested approaches. This can be partly explained by the low sequence variability in the region of the 16S rRNA gene, especially after read trimming to equal length, and the presence of closely related species and strains. In addition, all tested approaches generated more OTUs/ASVs than species in the dataset, exaggerating the species richness in the sample with many false positives, even after removal of chimeric sequences. This problem, also known as OTU inflation, forms a true challenge for clustering and most denoising algorithms and has been repeatedly reported [7, 62].

UPARSE, DADA2, and DGC demonstrated the highest accuracy, each correctly assigning a single ASV/OTU per reference sequence—consistent with prior pipeline comparisons [63]. In contrast, MED and UNOISE3 performed poorly, with <25% of reference sequences uniquely represented by a single OTU/ASV. Clustering-based algorithms frequently over-merged similar sequences (5–10% of OTUs), particularly with less stringent cutoffs [Additional file 7]. Notably, ASV-based methods like DADA2 and Deblur also exhibited over-merging (11% and 1%, respectively), likely due to

erroneous sequence binning where true variants were misclassified as noise and merged into unrelated ASVs.

Validation against 97%-clustered reference sequences revealed that OTU-based algorithms recovered all expected references, whereas DADA2 missed 3% (single-end) and 10% (paired-end) of references. For paired-end data, this may reflect failed merging of heavily erroneous forward and reverse reads, while single-end losses suggest over-aggressive error filtering. DADA2 with the default parameters reduced over-merging but introduced trade-offs: increased over-splitting and fewer correctly identified ASVs. Unlike the unified parameters, which enforce strict criteria across all algorithms, the default parameters risked both discarding true biological sequences and retaining more artifacts. Collectively, ASV methods showed higher over-splitting rates than OTU approaches, indicating that over-merging and over-splitting coexist as challenges in ASV-based pipelines.

When assessing the reflection of the approaches' varying performances on the microbial diversity, DADA2 and UPARSE (also Deblur to a lesser extent) achieved the closest resemblance -through sequence mapping- to the theoretical reference, aligning with findings from earlier studies [63–65]. This was reflected in the diversity analysis, where DADA2 and Deblur achieved the closest results to the theoretical reference. Which was observed by alpha and beta indices considering the microbial memberships solely or coupled with their abundances. However, DADA2 exhibited bacterial strain over-merging, consistent with findings of other studies [37, 66, 67]. UNOISE3 and MED on the other hand showed drastically different results from the intended theoretical diversity. For OTU-based approaches, UPARSE achieved the closest resemblance to the intended input, followed by DGC, which was manifested by both alpha and beta diversities and consistent with other pipeline comparative studies [67].

Despite the apparent differences in complexity, coverage, and the targeted 16S rRNA gene region between HC227\_V3V4 and the Mockrobiota mocks, there were no discrepancies in the overall comparison conclusions. Despite our attempts to unify sequencing depth and analyze multiple depths, biases introduced by sequencing methods still pose challenges that need to be acknowledged. Furthermore, our analysis exclusively utilized Illumina data targeting the V4/V3-V4 regions of the 16S rRNA gene, yet it is still unclear whether these results will differ when using different 16S rRNA regions (V1-3, V7-9, or V4-5), long-read sequencers (e.g. PacBio) or other amplicons (e.g. ITS) [34, 68]. Thus, the robust design of our benchmark analysis, combined with the use of the complex mock community data presented, provides a strong foundation for evaluating these

algorithms—a framework that could be extended to full pipeline comparisons in future studies.

## Conclusions

In conclusion, both OTU and ASV approaches produced varying results with pros and cons for each approach. ASV algorithms provided consistent sequence variants suitable for independent samples or meta-analysis studies without requiring re-clustering. DADA2 performed best among ASV approaches, preserving original diversity with little increase in over-merging, though its default parameters led to significantly higher over-splitting and mismatches. Conversely, for OTU-based algorithms, UPARSE performed the best, with balanced merging/splitting rates capable of handling inflated sequencing error/artifacts. The latter is suitable for under-examined niches or when a major microbial shift is to be expected. DADA2 and UPARSE provided the closest resemblance to the intended microbial community with the lowest error rate and the least number of artifacts.

## Abbreviations

OTU	Operational Taxonomic Units
ASV	Amplicon Sequence Variant
ESV	Exact Sequence Variants (ESVs)
zOTU	Zero Radius Operational Taxonomical Unit
SE	Single-End
PE	Paired-End
DADA2	Divisive Amplicon Denoising Algorithm 2
MED	Minimum entropy decomposition
AN	Average neighbor
DGC	Distance-based greedy clustering

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40793-025-00705-6>.

Supplementary Material 1: fig. 1: Re-analysis of DADA2 with default parameters. DADA2 was re-analyzed to remove the confounding effect of unified preprocessing and was subsequently compared to the results of our unified preprocessing (DADA2). (A) Stacked bar plots showing specificity analysis, representing the number of non-chimeric ASVs for each condition in both the Mockrobiota and HC227\_V3V4 mock communities. (B) Comparison of error rates between the unified preprocessing method (DADA2) and the default parameter method (DADA2\*) for the Mockrobiota and HC227\_V3V4 mock communities. (C) Box plots showing the difference in specificity analysis results between the unified preprocessing method (DADA2) and the default parameter method (DADA2\*) for the Mockrobiota mock community. (D) Box plots showing the results of Merging/Splitting Analysis between the unified preprocessing method (DADA2) and the default parameter approach (DADA2\*) for the Mockrobiota mock community. (E) Stacked bar plots illustrating the merging/splitting results compared to ASV-reference data for both the unified preprocessing method (DADA2) and the default parameter method (DADA2\*) for the HC227\_V3V4 mock community.

Supplementary Material 2: fig. 2: rarefaction curves and alluvial plots for both single-end and paired-end data. Rarefaction curves were illustrated for representing sequencing depth for denoising tools against *ASV-ref* and clustering methods against *OTU-ref* for both single-end and paired-end conditions respectively. Alluvial plots were utilized for visualizing the bacterial content across denoising and clustering algorithms for paired-end method.

Supplementary Material 3: fig. 3: illustration of HC227\_V3V4 utilizing

euclidian distance and alluvial plots for visualizing bacterial content for denoising and clustering methods. (A) Beta Diversity for denoising algorithms against *ASV-ref* data for both single-end and paired-end methods. (B) Beta Diversity for clustering algorithms against *OTU-ref* for both single-end and paired-end methods. C & D) stacked bar plots illustrating read count for all algorithms for Mockrobiota and HC227\_V3V4 mock communities considering single-end and paired-end conditions. E) Alluvial plots were utilized for visualizing the bacterial content across denoising and clustering algorithms for single-end method.

Additional file 1: representing details about Tool characteristics comparison.

Additional file 2: represents the available mock communities with their strain number to be used in benchmarking.

Additional file 3: represents the Mockrobiota data collected from web repository of microbial mock data.

Additional file 4: Describe the parameter used for running each algorithm.

Additional file 5: Wilcoxon sum rank statistical analysis results with pairwise comparison.

Additional file 6: Theoretical and expected 16S rRNA reference HC227\_V3V4.

Additional file 7: Parameter effect results on clustering-based algorithms.

Additional file 8: distance-based comparison across all clustering/denoising algorithms.

Additional file 9: Computational cost analysis results.

## Author contributions

Conceptualization: M.M, M.E, experimental design: M.M, P.V., R.V.H, lab work: I.C. results analysis: M.F, E.K.T, result Interpretation: M.F, M.M, M.E, P.V., R.V.H, P.M visualization: E.K.T, M.F. All authors contributed to writing and reviewing the paper.

## Funding

Not applicable.

## Data availability

16S rRNA amplicon sequence data generated and analyzed in this study have been deposited in the NCBI Sequence Read Archive with the accession code PRJNA975486.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

## Author details

<sup>1</sup>Bioinformatics Group, Center for Informatics Science, School of Information Technology and Computer Science, Nile University, Giza, Egypt

<sup>2</sup>Veterinary Research Institute, National Research Centre, Giza, Egypt

<sup>3</sup>Laboratory of Microbiology, Department of Biochemistry and Microbiology, Faculty of Sciences, BCCM/LMG Bacteria Collection, Ghent University, Ghent, Belgium

<sup>4</sup>Department of Biomedical Sciences, Institute of Tropical Medicine Antwerp, Antwerp, Belgium

<sup>5</sup>Microbiology Unit, Nuclear Medical Applications, Belgian Nuclear Research Centre (SCK CEN), Mol, Belgium

<sup>6</sup>Laboratory for Microbiology, Department of Biochemistry and Microbiology, Ghent University, Ghent, Belgium



Received: 13 March 2024 / Accepted: 13 April 2025

Published online: 13 May 2025

## References

1. Morgan XC, Huttenhower C. Chap. 12: human Microbiome analysis. *PLoS Comput Biol*. 2012. <https://doi.org/10.1371/journal.pcbi.1002808>.
2. Rausch P, Rühlemann M, Hermes BM et al. (2019) Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms. 1–19.
3. Mysara M, Leys N, Raes J, Monsieus P. IPED: A highly efficient denoising tool for illumina miseq Paired-end 16S rRNA gene amplicon sequencing data. *BMC Bioinformatics*. 2016. <https://doi.org/10.1186/s12859-016-1061-2>.
4. Mysara M, Saey Y, Leys N, Raes J, Monsieus P. CATCh, an ensemble classifier for chimera detection in 16s rRNA sequencing studies. *Appl Environ Microbiol*. 2015;81:1573–84.
5. Mysara M, Leys N, Raes J, Monsieus P. NoDe: A fast error-correction algorithm for pyrosequencing amplicon reads. *BMC Bioinformatics*. 2015. <https://doi.org/10.1186/s12859-015-0520-5>.
6. Kunin V, Engelbrekton A, Ochman H, Hugenholtz P. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ Microbiol*. 2010;12:118–23.
7. Reeder J, Knight R. The rare biosphere: A reality check. *Nat Methods*. 2009;6:636–7.
8. Caporaso JG, Kuczynski J, Stombaugh J, et al. Correspondence QIIME allows analysis of high-throughput community sequencing data intensity normalization improves color calling in solid sequencing. *Nat Publishing Group*. 2010;7:335–6.
9. Schloss PD, Westcott SL, Ryabin T, et al. Introducing Mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol*. 2009;75:7537–41.
10. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods*. 2013;10:996–8.
11. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: A versatile open source tool for metagenomics. *PeerJ*. 2016;2016:1–22.
12. Westcott SL, Schloss PD. (2017) OptiClust, an Improved Method for Assigning Amplicon-Based Sequence Data to Operational Taxonomic Units. *mSphere*. <https://doi.org/10.1128/mSphereDirect.00073-17>.
13. Edgar RC. Updating the 97% identity threshold for 16S ribosomal RNA OTUs. *Bioinformatics*. 2018;34:2371–5.
14. Mysara M, Vandamme P, Props R, Kerckhof FM, Leys N, Boon N, Raes J, Monsieus P. (2017) Reconciliation between operational taxonomic units and species boundaries. *FEMS Microbiol Ecol*. <https://doi.org/10.1093/femsec/fix029>.
15. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J*. 2017;11:2639–43.
16. Bandekar M, More KD, Seleyi SC, Ramaiah N, Kekäläinen J, Akkanen J. Comparative analysis of Microbiome inhabiting oxygenated and deoxygenated habitats using V3 and V6 metabarcoding of 16S rRNA gene. *Mar Environ Res*. 2024. <https://doi.org/10.1016/j.marenvres.2024.106615>.
17. Cao Z, Wang D, Hu X, He J, Liu Y, Liu W, Zhan J, Bao Z, Guo C, Xu Y. Comparison and association of winter diets and gut microbiota using TrnL and 16S rRNA gene sequencing for three herbivores in Taohongling, China. *Glob Ecol Conserv*. 2024. <https://doi.org/10.1016/j.gecco.2024.e03041>.
18. Li XM, Lv Q, Chen YJ, Yan LB, Xiong X. Association between childhood obesity and gut microbiota: 16S rRNA gene sequencing-based cohort study. *World J Gastroenterol*. 2024;30:2249–57.
19. Merali N, Chouari T, Sweeney C, et al. The microbial composition of pancreatic ductal adenocarcinoma: A systematic review of 16S rRNA gene sequencing. *Int J Surg*. 2024. <https://doi.org/10.1097/js9.0000000000001762>.
20. Osbelt L, Almási ÉdH, Wende M, et al. Klebsiella oxytoca inhibits Salmonella infection through multiple microbiota-context-dependent mechanisms. *Nat Microbiol*. 2024;9:1792–811.
21. Steimle A, Neumann M, Grant ET, et al. Gut microbial factors predict disease severity in a mouse model of multiple sclerosis. *Nat Microbiol*. 2024. <https://doi.org/10.1038/s41564-024-01761-3>.
22. Sun H, Chen F, Zheng W, Huang Y, Peng H, Hao H, Wang KJ. Impact of captivity and natural habitats on gut Microbiome in *Epinephelus akaara* across seasons. *BMC Microbiol*. 2024. <https://doi.org/10.1186/s12866-024-03398-y>.
23. Sun P-F, Lu MR, Liu Y-C, et al. An acidophilic fungus promotes prey digestion in a carnivorous plant. *Nat Microbiol*. 2024. <https://doi.org/10.1038/s41564-024-01766-y>.
24. Tang Q, Huang H, Xu H, Xia H, Zhang C, Ye D, Bi F. Endogenous Coriobacteriaceae enriched by a high-fat diet promotes colorectal tumorigenesis through the CPT1A-ERK axis. *NPJ Biofilms Microbiomes*. 2024. <https://doi.org/10.1038/s41522-023-00472-7>.
25. Callahan BJ, Wong J, Heiner C, Oh S, Theriot CM, Gulati AS, McGill SK, Dougherty MK. High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. *Nucleic Acids Res*. 2019;47:E103.
26. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from illumina amplicon data. *Nat Methods*. 2016;13:581–3.
27. Eren AM, Morrison HG, Lescault PJ, Reveillaud J, Vineis JH, Sogin ML. Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J*. 2015;9:968–79.
28. Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG, Sogin ML. Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol Evol*. 2013;4:1111–9.
29. Amir A, Daniel M, Navas-Molina J, et al. Deblur rapidly resolves single-. *Am Soc Microbiol*. 2017;2:1–7.
30. Plummer E, Twin J, Bulach DM, Garland SM, Tabrizi SN. A comparison of three bioinformatics pipelines for the analysis of preterm gut microbiota using 16S rRNA gene sequencing data. *J Proteom Bioinform Cit*. 2015;8:283–91.
31. D'Argenio V, Casaburi G, Precone V, Salvatore F. Comparative metagenomic analysis of human gut Microbiome composition using two different bioinformatic pipelines. *Biomed Res Int*. 2014;2014:325340.
32. Caruso V, Song X, Asquith M, Karstens L. Performance of Microbiome sequence inference methods in environments with varying biomass. *mSystems*. 2019;4. <https://doi.org/10.1128/mSystems.00163-18>.
33. Sirichoat A, Sankuntaw N, Engchanil C, Buppasiri P, Faksri K, Namwat W, Chantatit W, Lulitanond V. Comparison of different hypervariable regions of 16S rRNA for taxonomic profiling of vaginal microbiota using next-generation sequencing. *Arch Microbiol*. 2021;203:1159–66.
34. Abellan-Schneyder I, Machado MS, Reitmeier S, Sommer A, Sewald Z, Baumbach J, List M, Neuhaus K. Primer, pipelines, parameters: issues in 16S rRNA gene sequencing. *mSphere*. 2021. <https://doi.org/10.1128/mSphere.01202-20>.
35. Marizzoni M, Gurry T, Provasi S, et al. Comparison of bioinformatics pipelines and operating systems for the analyses of 16S rRNA gene amplicon sequences in human fecal samples. *Front Microbiol*. 2020. <https://doi.org/10.3389/fmicb.2020.01262>.
36. Prodan A, Tremaroli V, Rolin H, Zwiderman AH, Nieuwdorp M, Levin E. Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. *PLoS ONE*. 2020. <https://doi.org/10.1371/journal.pone.0227434>.
37. Nearing JT, Douglas GM, Comeau AM, Langille MGI. Denoising the denoisers: an independent evaluation of Microbiome sequence error-correction approaches. *PeerJ*. 2018;6:e5364.
38. Bokulich NA, Rideout JR, Mercurio WG, Shiffer A, Wolfe B, Maurice CF, Dutton RJ, Turnbaugh PJ, Knight R, Caporaso JG. (2016) mockrobiota: a Public Resource for Microbiome Bioinformatics Benchmarking. *mSystems*. <https://doi.org/10.1128/mSystems.00062-16>.
39. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. *Appl Environ Microbiol*. 2013;79:5112–20.
40. Comeau AM, Douglas GM, Langille MGI. (2017) Microbiome helper: a custom and streamlined workflow for Microbiome research. *mSystems*. <https://doi.org/10.1128/mSystems.00127-16>.
41. Goussarov G, Claesen J, Mysara M, Cleenwerck I, Leys N, Vandamme P, Van Houdt R. Accurate prediction of metagenome-assembled genome completeness by MAGISTA, a random forest model built on alignment-free intra-bin statistics. *Environ Microbiomes*. 2022;17:1–13.
42. Kechin A, Boyarskikh U, Kel A, Filipenko M. CutPrimers: A new tool for accurate cutting of primers from reads of targeted next generation sequencing. *J Comput Biol*. 2017;24:1138–43.
43. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26:2460–1.
44. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*. 2011;27:863–4.



45. Sasada R, Weinstein M, Prem A, Jin M, Bhasin J. FIGARO: an efficient and objective tool for optimizing Microbiome rRNA gene trimming parameters. *J Biomol Tech*. 2020;31:S2.
46. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Priesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO. The SILVA and all-species living tree project (LTP) taxonomic frameworks. *Nucleic Acids Res*. 2014;42:643–8.
47. Quast C, Priesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2013;41:590–6.
48. Edgar RC. (2016) UNOISE2: improved error-correction for illumina 16S and ITS amplicon sequencing. *bioRxiv* 081257.
49. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–10.
50. Florens L, Hao Y, Zhang Y, Wen Z, Banks CAS, Washburn MP. (2020) Poster abstracts abrf 2020. 31:2020.
51. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J*. 1948;27:379–423.
52. Brose U, Martinez ND. Estimating the richness of species with variable mobility. *Oikos*. 2004;105:292–300.
53. Jaccard P. THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE. *New Phytol*. 1912;11:37–50.
54. Gower JC, Legendre P. (1986) Metric and Euclidean Properties of Dissimilarity Coefficients.
55. Lance GN, Williams WT. Computer programs for hierarchical polythetic classification (similarity analyses).
56. Lance GN, Williams WT. Mixed-Data classificatory programs I - Agglomerative systems. *Aust Comput J*. 1967;1:15–20.
57. Pedersen TL. (2024) patchwork: The Composer of Plots.
58. Wickham H. *ggplot2: elegant graphics for data analysis*. New York: Springer; 2016.
59. Brunson JC, Read QD. (2023) ggalluvial: Alluvial Plots in ggplot2.
60. Wickham H, Vaughan D, Girlich M. (2024) tidy: Tidy Messy Data.
61. Reitmeier S, Hitch TCA, Treichel N, et al. Handling of spurious sequences affects the outcome of high-throughput 16S rRNA gene amplicon profiling. *ISME Commun*. 2021. <https://doi.org/10.1038/s43705-021-00033-z>.
62. Edgar RC. Accuracy of microbial community diversity estimated by closed- and open-reference OTUs. *PeerJ*. 2017. <https://doi.org/10.7717/peerj.3889>.
63. Jeske JT, Gallert C. Microbiome analysis via OTU and ASV-Based Pipelines—A comparative interpretation of ecological data in WWTP systems. *Bioengineering*. 2022. <https://doi.org/10.3390/bioengineering9040146>.
64. Glassman SI, Martiny JBH. BROADSCALE ecological patterns are robust to use of exact. *mSphere*. 2018;3:e00148–18.
65. García-López R, Cornejo-granados F, Lopez-zavala AA, Cota-Huizar A, Sotelo-Mundo RR, Gómez-Gil B, Ochoa-Leyva A. OTUs and ASVs produce comparable taxonomic and diversity using tailored abundance filters. *Genes (Basel)*. 2021;12:564.
66. Odom AR, Faits T, Castro-Nallar E, Crandall KA, Johnson WE. Metagenomic profiling pipelines improve taxonomic classification for 16S amplicon sequencing data. *Sci Rep*. 2023;13:1–12.
67. Allali I, Arnold JW, Roach J, et al. A comparison of sequencing platforms and bioinformatics pipelines for compositional analysis of the gut Microbiome. *BMC Microbiol*. 2017. <https://doi.org/10.1186/s12866-017-1101-8>.
68. Fadeev E, Cardozo-Mino MG, Rapp JZ, Bienhold C, Salter I, Salman-Carvalho V, Molari M, Tegetmeyer HE, Buttigieg PL, Boetius A. Comparison of two 16S rRNA primers (V3–V4 and V4–V5) for studies of Arctic microbial communities. *Front Microbiol*. 2021;12:1–11.

## Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.